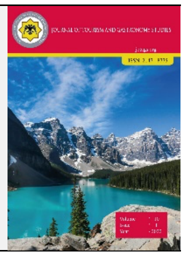
JOURNAL OF TOURISM AND GASTRONOMY STUDIES

ISSN: 2147 – 8775

Journal homepage: www.jotags.org



Analysis of Fine Dining Restaurant Reviews for Perception of Customers' Restaurant Service Quality

*** Semra AKTAS POLAT** [a]

[a] Istanbul Medeniyet University, Faculty of Tourism, Department of Tourism Management, Istanbul/Turkey

Article History

Received: 12.02.2022

Accepted: 15.03.2022

Keywords

Perceived service quality

Online reviews

Fine dining restaurants

Aspect-based sentiment analysis

Latent Dirichlet allocation

Machine learning

Article Type

Research Article

Abstract

The purpose of this study is to model the perception of customers' service quality in fine dining restaurants (FDRs) and to determine customer sentiments towards the service quality. I analyzed 22,104 reviews of 25 restaurants on TripAdvisor through Aspect-Based Sentiment Analysis (ABSA). In terms of n-gram language models, the classification performance of sentiment polarity was tested with Support Vector Machine (SVM), Naive Bayes (NB), C4.5, and Gradient Boosted Trees (GBT). I compared the performance of the model with Cohen's kappa, accuracy, precision, recall, and F-measure results. I found five topic models service, experience, surprise, taste, and food kind by using latent Dirichlet allocation (LDA). In sentiment classification, SVM achieved the best results in bigram with 74.5% average F-measure, 94.4% accuracy, and 49.2% kappa results. This study contributes to the elements related to the perception of service quality in FDRs with psychological quality proposed by the surprise topic. This is one of the few studies conducted with ABSA on the perception of service quality in FDRs, and it is the first study examining the issue in terms of n-gram language models.

** Corresponding Author*
E-mail: saktaspolat@gmail.com (S. Aktaş Polat)

DOI:10.21325/jotags.2022.974

INTRODUCTION

Restaurant service quality provides emotional comfort to customers (Johns & Howard, 1998) with indicators such as speed of service, preperation, menu (Huang, 2003), staff, equipment, appearance, and individual interest towards guests (Lisnawati & Astawa, 2020). In this context, service quality, one of the basic components of restaurant experience (Gagić et al., 2013), is defined as one of the restaurant characteristics with elements such as theme concept, food quality, menu, and atmosphere (MacLaurin & MacLaurin, 2000). Moreover, providing service quality is a critical factor in creating competitive advantage in the restaurant industry (Jin et al., 2013; Madanoglu, 2006). Therefore, service quality is crucial for the success of a restaurant (Keith & Simmers, 2011) and determining the food satisfaction of consumers (Lee et al., 2016).

Consumers' judgments about the superiority of service refer to the quality of service (Kala, 2020), and the judgments about the general excellence or superiority of a business refer to perceived quality (Zeithaml, 1987). The perceived quality of service is a result of the comparison of consumers with the expected and perceived service (Grönroos, 1984; Parasuraman et al., 1985). Managing service quality is about providing service by designing the service product and service environment (Rust & Oliver, 1994). However, consumers' quality definition has become more important than managements' definition (Berry et al., 1988). Knowing how service quality is perceived from a customer perspective will help businesses achieve some goals including improving product or service quality (Schuckert et al., 2015), positive Word-of-Mouth, and brand loyalty (Huang, 2003). In the absence of objective factors, consumers' quality perceptions are measured to evaluate a firm's service quality (Parasuraman et al., 1988). Many studies (Antun et al., 2010; Bitner, 1992; Cronin & Taylor, 1992; Grönroos, 1984; Parasuraman et al., 1988; Raajpoot, 2002; Ryu & Jang, 2008; Stevens et al., 1995) have been conducted to measure the perception of service quality. The main purpose of service quality measurement is to reveal what customers think about services they experience (Hansen, 2014).

Although the perception of service quality varies according to the characteristics of customers and restaurants (Kim et al., 2003), I can state that the perception of quality will be a holistic summary of the restaurant experience. Nevertheless, the intangible nature of services makes it difficult for consumers to evaluate the quality of service (Bojanic & Rosen, 1994). According to Wall & Berry (2007), customers benefit from three clues to evaluate the restaurant experience; functional, mechanic, and humanic.

The widespread use of the internet, especially Web 2.0, has made it easier for customers to share their thoughts about any product or service they experience. In the hospitality industry, online reviews have become an important source of information in potential customers' decision-making process (Pang & Lee, 2008). Until the advent of online reviews, it was difficult to reach customers in terms of cost and time to collect the data needed to measure service quality with surveys as the primary data collection tool (Palese & Piccoli, 2016). Unmeasurable amounts of food-related data is generated worldwide through online posts (Tao et al., 2020). However, the fact that this data turn into a pile over time makes it difficult to make sense. Hence, businesses may have difficulties in understanding how customers perceive the quality of service (Parasuraman et al., 1985). In recent years, data mining transforming raw data into useful information (Tan et al., 2014) and text mining, a variation of data mining (Hearst, 2003), have attracted attention to overcome these difficulties. While many industries place great emphasis on analyzing big data, the hospitality industry has not paid enough attention to the issue (Kim et al., 2017).

The use of data mining in the hospitality industry will be valuable, as it contains large amounts of information about people's movements and activities (Bermingham & Lee, 2014). Online reviews provide an open forum reflecting customers' current preferences and identify service dimensions that positively (or negatively) affect the overall service experience (Korfiatis et al., 2018). Moreover, increasing user-generated content provides opportunities for enterprises to monitor service quality (Alaei et al., 2019) and new approaches to service quality measurement (Palese & Usai, 2018). In this study, I used sentiment analysis, one of the text mining methods. I focused on the perception of service quality in fine dining restaurants (FDRs) through online reviews posted by customers with the idea that more research is needed on the perception of customers' service quality (Grönroos, 1984).

Literature review

Restaurant Service Quality

Grönroos (1984) claimed that consumers are interested in not only what they receive but also with the process itself and defined two types of quality for service quality technical and functional. Parasuraman et al. (1988) developed SERVQUAL having five dimensions reliability, empathy, responsiveness, tangibles, and assurance for customers' service expectations. Cronin and Taylor (1992) developed SERVPERF, a performance-based measurement based on SERVQUAL, by claiming that service quality should be measured as an attitude.

Bitner (1992) focused on the environmental dimensions of the service area with servicescapes under three headings ambient conditions, spatial layout and functionality, and artifacts, symbols, and signs. DINESERV (Stevens et al., 1995), TANGSERV (Raajpoot, 2002), DINESCAPE (Ryu & Jang, 2008), and DinEx (Antun et al., 2010) are the other leading attempts to measure restaurant service quality. Since it is used as a general tool to measure the gap between customer expectations and perceptions of service quality (Knutson et al., 1996), most of the tools used in the studies on restaurant services include five dimensions of SERVQUAL (Keith & Simmers, 2011).

Stevens et al. (1995) adapted the SERVQUAL instrument to the restaurant industry and developed the five-dimensional (tangibles, reliability, responsiveness, assurance, and empathy) DINESERV.

Raajpoot (2002) focused on the tangibles from the dimensions of service quality of SERVQUAL and DINESERV, by claiming that previous studies have completely ignored the ambient dimension. In this perspective, Raajpoot (2002) developed TANGSERV that consists of three dimensions; layout, product/service, and ambiance/social to measure tangible quality in the food service industry. Furthermore, Ryu & Jang (2008) investigated the dimensions of upscale restaurants' physical environment, excluding outdoor environments and interior areas where no food is served, and they developed the DINESCAPE scale consisting of facility layout, ambience, social factors, lighting, aesthetics, and service product.

Antun et al. (2010) developed DinEx, which defines the expectations of restaurant guests. The DinEx scale shows that restaurant guests have social and health expectations as well as food, service, and atmosphere expectations. Nevertheless, customer expectations may change; for example, Marković et al. (2010) found that the highest expectation scores for the restaurant service quality are the elements that fall under the dimensions of reliability and tangibles. Moreover, Johns & Tyas (1996) found that food and staff attitudes are more important in the catering industry. Knutson et al. (1996) determined that customers' expectations are reliability, tangibles, assurance, responsiveness, and empathy in fine dining, casual, and quick service restaurants.

Since it is a critical factor for businesses, the need for new studies to measure service quality continues despite all these attempts (Mejia et al., 2020). Measurement and quantification difficulties in the service industry have made it difficult for managers to monitor and control processes (Chase & Apte, 2007). In order to overcome this difficulty, the use of online customer feedback has become widespread in addition to traditional approaches in service quality analyses (James et al., 2017). However, quantitative interpretation of user-generated content alone is not sufficient for a comprehensive and accurate assessment (Duan et al., 2016).

I think that the sentiments describing customers' experiences are crucial in revealing the dimensions of service quality in FDRs with the idea that experience is more important than meeting the hunger need because of the fact that customers accept a reservation period that takes weeks even months to have the dining experience at these fine dining restaurants. In addition to customer thoughts, suggestions, recommendations, and complaints I paid more attention to customer sentiments through the online reviews posted by customers on TripAdvisor in order to reveal whether there is a service quality dimension different from the previous studies in terms of the perception of service quality in the FDRs experience. I preferred to use online reviews due to online reviews can be more objective for the measurement of restaurant service quality (Duan et al., 2016) and online reviews contain customers' own sentences instead of structured statements. For this reason, I aim to be able to interpret customers' fine dining experiences by examining their own sentences.

Sentiment Analysis on Restaurant Reviews

Sentiment analysis aims to determine the perspectives underlying a text range (Pang & Lee, 2004). Sentiment analysis is used in many fields as a promising method, but it has not been used much in hospitality and tourism (Philander & Zhong, 2016). However, interest has been increasing in tourism studies (Kirilenko et al., 2018). Fu et al. (2019) claimed that sentiment is used as a synonym for emotion and attitude in tourism studies. According to Li et al. (2020), emotional expressions in online reviews can be based on internal (e.g., the characteristics of reviewer) or external factors (e.g., service quality). In the current literature, it is seen that online reviews were directly examined within the scope of service quality in the fields of accommodation (Duan et al., 2016; Moro et al., 2020), airlines (Lim & Lee, 2020), and restaurants (Mejia et al., 2020).

Duan et al. (2016) used sentiment analysis to divide online customer reviews into the five dimensions of SERVPERF by measuring the hotel service quality, and they found that the size of tangibles had the highest number of sentences with 69.66%. Moro et al. (2020) analyzed TripAdvisor reviews for the service quality of a high-end and a low-end chain airport hotel operating in five different cities of Europe with latent Dirichlet allocation (LDA). The authors identified seven important dimensions staff, reservation, cleanliness, transportation, value, schedule, and food and beverage.

Lim & Lee (2020) analyzed online reviews written by passengers for airline services and they found that while the most important dimensions were tangibles and reliability, the least important dimensions were assurance and empathy. Mejia et al. (2020) analyzed restaurant reviews and determined five components, namely overall quality, wait times, food quality, responsiveness, and atmosphere, which reflect the characteristics of a restaurant, by using the nonnegative matrix factorization technique, one of the topic modeling techniques.

In addition to the studies mentioned above, there are also some studies examining restaurant reviews on a technical basis. Kang et al. (2012), for example, created a dictionary of sentiment classification containing unigrams and bigrams from restaurant reviews. In addition, Zhao et al. (2016) developed the Service Quality Evaluation Model algorithm to assess service quality through the user ratings in Yelp (restaurants and nightlife) and Duban (movie reviews) datasets. Moreover, Akhtar et al. (2017) suggested Maximum Entropy (ME), Conditional Random Field (CRF), and SVM to classify restaurant and laptop reviews with the Particle Swarm Optimization (PSO) method they proposed for Aspect Based Sentiment Analysis (ABSA). Furthermore, García-Pablos et al. (2018) used the reviews on hotels, restaurants, and electronic devices in different languages (English, Spanish, French, and Dutch) for a method proposal called W2VLDA based on topic modeling for domain aspect and sentiment classification in ABSA.

Although the main purpose of some studies examining restaurant reviews is not to determine the perception of service quality, they have some findings related to the perception of service quality including satisfaction (Aktas-Polat & Polat, 2022; Geler et al., 2021; Pantelidis, 2010), intention to revisit (Yan et al., 2015), restaurant aspects extraction (Luo & Xu, 2019), the factors affecting the sentiment towards dining out (Tian et al., 2021), and customer value (Kwon et al., 2020).

Determining the perception of service quality is a key issue for the restaurant industry. Especially in fine-dining restaurants, it is crucial to determine service quality characteristics, since customers pay attention to service quality as well as providing quality meals (Cheng et al., 2012). This study differs in terms of the method used by analyzing online restaurant reviews with ABSA and LDA. Furthermore, this study used C4.5 and Gradient Boosted Trees (GBT) unlike the studies that used NB and SVM for sentiment classification performance in restaurant sample (Akhtar et al., 2017; Kang et al., 2012; Luo and Xu, 2019).

Online reviews are widely accepted due to their up-to-dateness and sample size (Luo et al., 2021). This study handled online reviews as a source of information on service quality (Mejia et al., 2020) and focused on the perception of service quality in FDRs. Despite the current studies, I wonder whether there are service quality dimensions to be revealed by machine learning algorithms in online customer reviews. The study is trying to answer the following four research questions (RQ):

RQ1: What are the topic models for the perception of service quality in FDRs?

RQ2: What is the customer sentiment polarity towards the topic models of service quality in FDRs?

RQ3: What is the best performing n-gram language model in the sentiment classification?

RQ4: What is the best performing supervised machine learning algorithm in the sentiment classification?

Material and Methods

The aim of this study is to model the perception of service quality in FDRs and to determine customer sentiments towards service quality. Sentiment analysis can be performed in three different ways; at document level, sentence level, and on an aspect basis (Feldman, 2013). Aspect Basis Sentiment Analysis (ABSA) used in this paper focuses on the features or functions of products (Zhang et al., 2012). With ABSA, the entities and their aspects are initially identified and extracted in documents, then the sentiment polarities of these entities and aspects are determined (Zhang & Liu, 2014). In this study, FDRs have been considered as the entity and focused on the second and third sub-tasks. Latent Dirichlet allocation, one of the topic modeling methods, was used to determine the basic aspects of the entity (Bagheri et al., 2014; Luo & Xu, 2019). Figure 1 shows the stages followed in the study.

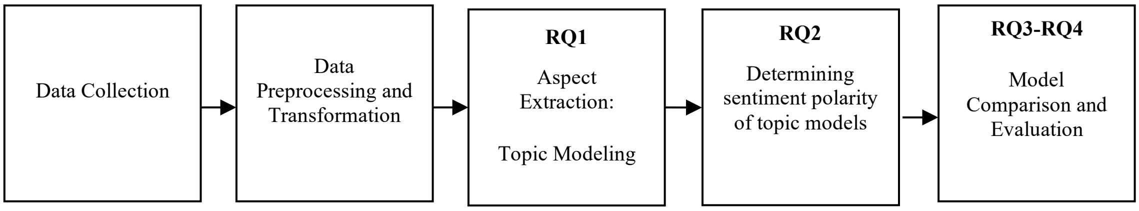


Figure 1. The Stages of the Study

Data Collection and Preprocessing

I analyzed the customer reviews with topic modeling due to the existence of quality dimensions that could not be directly obtained from the current review interface of TripAdvisor and therefore not measured (Korfiatis et al., 2018). I obtained the data used in the study from the list of Best Fine Dining Restaurants-World published by TripAdvisor (2021). The dataset of the study consists of 22,104 manually collected reviews in English between 2004 and 2021 for 25 restaurants operating in 16 countries. All transactions in the study were done with Knime Analytics Platform 4.3.1.

Aspect Extraction: Topic Modeling with LDA

Although the experiences of customers are different, the words they use combine when they assess a product's features (Hu & Liu, 2004). The detection of this association is important for feature detection expressed as the second task of ABSA. For this task, I applied the LDA algorithm in the study. The number of topics is an important element in topic modeling (Sutherland et al., 2020). Elbow method is used to determine the most suitable number of topics for LDA (Aktas-Polat & Polat, 2022; Taecharungroj & Mathayomchan, 2019).

Determining Aspect Sentiment Polarity

Classifying a text containing opinions as positive and negative is called polarity classification (Pang & Lee, 2008). In this paper, customers' ratings were used to determine sentiment polarity (Pang et al., 2002), and while 4 and 5-star ratings are labeled positive, 1, 2, and 3-star ratings are labeled as negative (Aktas-Polat & Polat, 2022; Taecharungroj & Mathayomchan, 2019).

Comparison of Model Performances of Supervised Machine Learning Algorithms

In this study, the performance of sentiment classification was tested on the base of n-gram models. The most common practice for n-gram, which refers to a word sequence that detects dependencies between words, is the use of unigram, bigram, and trigram (Yousefpour et al., 2014). The system needs, in n-gram models, to look at the previous n-1 words to predict the nth word (Bhuyan & Sarma, 2019). I have used unigram, bigram, trigram, and quadrigram from n-gram language models. According to n-gram language models, the performance of sentiment classification is tested with C4.5 developed for Decision Tree by Quinlan (1993) and GBT, one of the decision trees ensemble models, in addition to the NB algorithm (Kang et al., 2012; Luo & Xu, 2019) and the SVM algorithm used for comparison in restaurant datasets (Akhtar et al., 2017; Kang et al., 2012; Luo & Xu, 2019).

I used the widely used k-fold cross validation (k-cv) method (Bengio & Grandvalet, 2004) to estimate the prediction error in separating the data into training and test data. This method was run with the 10-fold cross validation. According to Dey et al. (2018), the performance of algorithms is compared with the evaluation parameters, namely Recall, Precision, F-measure, and Accuracy scores. Evaluating the results of sentiment analysis is complicated by the fact that Cohen's kappa is rarely used for performance evaluation (Kirilenko et al., 2018). Because of this I evaluated Cohen's kappa results for the performances of the algorithms in addition to other parameters. The kappa (κ) used as a measure of agreement is the interjudge agreement coefficient for nominal scales (Cohen, 1960). The lower limit of κ depending on the distribution of two reviewers' judgments is between 0 and -1.00 while the upper limit of it is +1.00 (Cohen, 1960). According to Landis and Koch (1977), the ranges of κ can be labeled as poor (<0.00), slight (0.00–0.20), fair (0.21–0.40), moderate (0.41–0.60), substantial (0.61–0.80), and almost perfect (0.81–1.00) in order to provide useful benchmarks for the relative strength of agreement. A value above 0.40 can be interpreted as adequate agreement (McHugh, 2012).

Results

Aspect Extraction: Topic Modeling for Service Quality

I used the elbow method to determine the optimal number of topics for LDA, and 5 was determined by the method as the elbow point. After this process, topic models related to the perception of service quality were determined by LDA for RQ1. LDA assigned 22,095 out of 22,104 reviews to a topic. Table I presents the first 10 words representing the topics generated by LDA with their weights.

Table I. Topic Models of Service Quality for FDRs

Service		**Experience**		**Surprise**		**Taste**		**Food Kind**	
Term	Weight	Term	Weight	Term	Weight	Term	Weight	Term	Weight
food	10216	food	13001	experience	10067	course	10265	dish	6819
restaurant	9772	restaurant	12248	food	8162	food	9573	course	3706
wine	7677	wine	11913	restaurant	4729	menu	8944	**dessert**	2756
service	7235	service	9156	course	3566	wine	7294	**fish**	2662
experience	4718	experience	7552	dish	3385	restaurant	6794	menu	2505
menu	4608	course	7187	meal	3337	staff	6788	meal	1681
course	4248	menu	6932	**amazing**	3324	experience	4653	restaurant	1667
meal	3345	meal	5042	service	3194	service	4296	**main**	1628
dish	2909	staff	4188	time	3071	meal	3657	**cream**	1626
excellent	2900	time	4084	staff	2979	**tasting**	3643	**bread**	1514

The more likely word in a topic model has more explanatory power (Lim & Lee, 2020). In labeling the topic models, therefore, manual content analysis was performed by reading these word groups and the top 10 reviews with the highest probability assigned to the relevant topic model (Guo et al., 2017; Sutherland et al., 2020). Moreover, in the labeling phase, I focused on the distinctive words (Taecharungroj et al., 2021) marked as bold in Table I, for I think they represent each topic model

Topic Model 1: Service

Excellent is the distinguishing word of this model. According to the content analysis conducted, this topic model emphasizes that service is the element that makes the experience unique. This emphasis is seen in the following reviews assigned to the service topic model "It was just the top notch service that made this truly a meal to remember" (id: 9982), and "The service is exemplary. Really, you feel so special eating there" (id: 18209).

Topic Model 2: Experience

There is no distinctive term for this model. According to the content analysis conducted, this model focuses on the FDR experience itself. For example, a customer defined this experience as "I suppose it's one of those things that people who love fine dining have to try once-like climbing Everest because it's there. It was definitely an experience" (id: 290). In this topic model, the experience is defined as artistic, experimental, and creative that make the person feel special.

Topic Model 3: Surprise

Amazing is the distinguishing term of this model. On the other hand, it was addressed that this topic model is an extension of the previous topic model experience as a result of content analysis. In particular, the terms experience, amazing, and time emphasize the content of FDRs for the surprises that are remembered. We can understand the importance of surprises for customers from the following reviews: "It is difficult to explain why we loved the restaurant so much without spoiling the surprises, I just think every foodie needs to experience the magic" (id: 10154), and "I have to say I wished I couldn't see what was happening on other tables as this did spoil the surprise slightly when it was our turn" (id: 43). From the reviews assigned to this topic model, I have seen that the reviewers expressed the confusion and emotional states they experienced because of the surprises they encountered during their experiences in FDRs. I claim that this topic model can be associated with the psychological output of service quality.

Topic Model 4: Taste

Tasting is the distinguishing word of this model. According to the content analysis conducted, it was seen that the fourth topic model is related to the taste of the dishes presented in FDRs. In particular, the terms course, food, menu, wine, meal, and tasting emphasize the taste of each element offered to the customers in FDRs. The following reviews, for example, show the importance that customers give to the taste of the dish they eat: "Hot bread rolls were brought out next, made with Black Sheep ale, very tasty. And then followed course upon course of some of the most delicious food I have ever eaten!" (id: 4922), and "The food-this is the reason for visiting and the menu, the tastes, the presentation the matching wines were all exemplary. ... I loved the black pudding. I was tempted by the smoked eel and scrambled eggs but resisted. ... Dislikes-none" (id: 10458).

Topic Model 5: Food Kind

The distinguishing terms of this model were found as dessert, fish, main, cream, and bread. According to the content analysis conducted, it was seen that the fifth topic model is related to the food kind. The following reviews show that customers emphasize the food kinds: "The mock turtle soup served next was another old dish, but today seemed less impressive than its previous version, technically clever but lacking depth of flavour" (id: 12762), and

"For desserts we ate the home made knaffe, creme brûlée, hel ice cream. We also had the anis grapefruit sorbet but it was a bit odd for me" (id: 16884).

Sentiment Polarity of Topic Models

I used customer star-rating in the sentiment classification of topic models regarding service quality for RQ2. Table II presents the distribution of positive and negative reviews in each topic model.

Table II. Sentiment Polarity of Topic Models

Topic Models	Review Count	Positive		Negative	
		n	%	*n*	%
Service	5182	4994	96.4	188	3.6
Experience	6820	6239	91.5	581	8.5
Surprise	3835	3579	93.3	256	6.7
Taste	4729	4395	92.9	334	7.1
Food Kind	1529	1337	87.4	192	12.6

According to Table II, while service is the topic model with the highest rate of positive sentiment, food kind is the topic model with the lowest rate. In parallel with this, service is the topic model with the lowest rate of negative sentiment while food kind is the topic model with the highest rate. Morevover, according to review count, experience (30.9%) was the most important topic model while food kind (6.9%) was the least important topic model.

Model Comparison and Evaluation

For RQ3 and RQ4, sentiment classification based on star ratings was tested with SVM, C4.5, NB, and GBT algorithms based on n-gram language models. Table III presents the performance results of the algorithms.

According to Table III, while the average values are between 57.3%–81% for recall, 60.8%–80% for precision, and 60%–70.5% for F-measure in unigram, they are between 58.7%–70.8% for recall, 60.1%–82.2% for precision, and 60%–74.5% for F-measure in bigram. Moreover, in trigram, the average values are between 57.9%–67.4% for recall, 59.5%–81.9% for precision, and 59.1%–71.5% for F-measure while they are between 57.2%–65.5% for recall, 60%–80.6% for precision, and 59.5%–70.1% for F-measure in quadrigram.

According to the average values, the highest performing algorithms were SVM, GBT, and SVM in unigram and bigram while NB, GBT, and SVM were in trigram, and NB and SVM, SVM, and SVM were in quadrigram. Nevertheless, in all the language models, GBT, C4.5, and C4.5 were the lowest performing algorithms for recall, precision, and F-measure respectively. In terms of the language models, the accuracy values are between 89.8%–94.1% for unigram, 86.8%–94.4% for bigram, 88.3%–94.3% for trigram, and 89.8%–94.1% for quadrigram. SVM provided the highest accuracy value in all the language models. Nevertheless, NB provided the lowest accuracy value in unigram, bigram, and trigram while C4.5 provided the lowest accuracy in quadrigram.

Table III. Model Comparison

Algorithm	Sentiment	N-gram									
		Unigram					Bigram				
		R	*P*	*F1*	*Accuracy*	*κ*	*R*	*P*	*F1*	*Accuracy*	*κ*
SVM	POS	95.1	98.8	96.9	94.1	41.4	98.3	95.8	97	94.4	49.2
	NEG	66.9	32.9	44.1			43.2	65.4	52		
	AVG	81	65.9	70.5			70.8	80.6	74.5		

Table III. Model Comparison (cont.)

	POS	95.3	94.3	94.8			94.5	94.4	94.4		
C4.5	NEG	23.3	27.4	25.2	90.3	20	25.5	25.8	25.6	89.6	20.1
	AVG.	59.3	60.8	60			60	60.1	60		
	POS	93.8	95.2	94.5			89.7	95.8	92.6		
NB	NEG	36.6	31	33.6	89.8	28.1	47.3	25.8	33.4	86.8	26.7
	AVG	65.2	63.1	64.1			68.5	60.8	63		
	POS	99..4	93.9	96.6			99.4	94.1	96.7		
GBT	NEG	15.1	65.9	24.6	93.5	22.5	17.9	70.2	28.6	93.7	26.5
	AVG	57.3	80	60.6			58.7	82.2	62.7		
Algorithm	**Sentiment**	**Trigram**					**Quadrigram**				
		R	***P***	***F1***	***Accuracy***	***κ***	***R***	***P***	***F1***	***Accuracy***	***κ***
	POS	98.8	95.3	97			98.7	95.1	96.9		
SVM	NEG	34.7	67.8	45.9	94.3	43.2	32.3	66	43.3	94.1	40.6
	AVG	66.75	81.6	71.5			65.5	80.6	70.1		
	POS	94.8	94.2	94.5			94.8	94.3	94.5		
C4.5	NEG	22.6	24.8	23.7	89.8	18.2	23.7	25.7	24.6	89.8	19.2
	AVG	58.7	59.5	59.1			59.3	60	59.5		
	POS	91.7	95.5	93.6			93.8	95.2	94.5		
NB	NEG	43.1	28.2	34.1	88.3	28	37.2	31.3	34	89.9	28.6
	AVG	67.4	61.9	63.9			65.5	63.3	64.3		
	POS	99.5	94	96.7			99.4	93.9	96.6		
GBT	NEG	16.2	69.7	26.3	93.6	24.3	15	67.1	24.5	93.5	22.5
	AVG	57.9	81.9	61.5			57.2	80.5	60.1		

Note. POS = Positive; NEG = Negative; AVG = Average

In terms of the language models, Cohen's kappa values are between 20%–41.4% for unigram, 20.1%–49.2% for bigram, 18.2%–43.2% for trigram, and 19.2%–40.6% for quadrigram. Based on the division proposed by Landis and Koch (1977), in terms of Cohen's kappa statistic, SVM can be expressed as the highest and C4.5 as the lowest aggrement algorithms in all language models. Furthermore, SVM was the only algorithm that performed above 0.40 which is the lowest value for adequate agreement (McHugh, 2012) or higher.

According to Table III, the best language model based on average precision, F-measure, accuracy and Cohen's kappa statistic was bigram followed by trigram, unigram, and quadrigram respectively. In bigram, SVM achieved the best results with an average F-measure of 74.5%, accuracy of 94.4%, and Cohen's kappa statistic of 49.2%. Moreover, GBT was the best algorithm in terms of 82.2% average precision. This situation can be explained by the fact that GBT was the algorithm with the lowest recall rate for negative. Nevertheless, SVM achieved the best performance with an average recall result of 81% in unigram followed by bigram, trigram, and quadrigram respectively. In terms of recall values, the highest value for negative was SVM with 66.9% in unigram while NB with 47.3% in bigram. The negative recall score of SVM showed a decrease by 23.7% in bigram compared to unigram, 8.5% in trigram compared to bigram, and 2.4% in quadrigram compared to trigram.

For RQ3, bigram was determined as the best performing n-gram language model for sentiment classification of online customer reviews for FDRs. However, in terms of sentiment classification, SVM was identified as the highest performing supervised machine learning algorithm for RQ4.

Discussion and Implications

In this study, I acted with the idea that although customer reviews are independent from each other, they are written around a common thought (Hu & Liu, 2004). I think that service quality in FDRs can be interpreted through online customer reviews. I found five topic models for service quality with LDA used for aspect extraction in ABSA.

In terms of SERVQUAL (Parasuraman et al., 1988) and DINESERV (Stevens et al., 1995), service is an extension of assurance and empathy; experience is an extension of all the service quality dimensions; surprise is an extension of responsiveness and empathy; and taste and food kind are the extensions of tangibles dimension.

While taste, one of the topic models determined in this study, overlaps with the food factor dimension of DinEX (Antun et al., 2010), food kind coincides with food variety, one of the dimensions of TANGSERV (Raajpoot, 2002). Moreover, service topic overlaps with DINESCAPE's service staff dimension, which refers to the employees who make customers feel good (Ryu & Jang, 2008), as well as DinEX's social factor and service factor dimensions (Antun et al., 2010). According to the functional and technical quality definition of Grönroos (1984), service, experience, and surprise can be interpreted as functional quality elements while taste and food kind as technical quality elements. Moreover, when I assess the topics of this study on the basis of the quality classification of Hwang and Ok (2013), I can express that service and experience indicate interactional quality; taste and food kind indicate outcome quality; and surprise indicates the intersection point of interaction and outcome qualities. Furthermore, in terms of Wall and Berry's (2007) functional, mechanic, and humanic clues, I can state that taste and food kind are functional; experience is mechanic and humanic; and service and surprise are humanic clues.

Kim et al. (2003) found that high average-spending diners expect individual attention. This study's topic models, especially service, experience, and surprise support them. In addition, I claim that experiencing unforgettable moments matters to FDR customers, and these moments will make customers feel good psychologically. Therefore, the recommendation of the study for surprise, one of the topic models derived from this study, was psychological quality. I can define psychological quality as the quality dimension of service that addresses a customer's mental state. Although psychological quality is similar to the emotional comfort category proposed by Johns and Howard (1998), it refers to psychological well-being beyond emotional comfort. Psychological quality is attributed some symbolic meanings such as status, dignity, prestige, and self-realization that respond to psychological needs through food (Aktas-Polat & Polat, 2020). Moreover, psychological quality refers to the psychological well-being accompanying the sense of delight experienced by customers with the surprises offered in FDRs. In addition, I can define surprises as key factors mediating the psychological well-being of customers. In this respect, I support Johns and Howard's (1998) assertion that the list has not yet been completed, and new service quality determinants are waiting to be discovered.

I also found that the most crucial topic model for customers in FDRs was experience followed by service, taste, surprise, and food kind respectively. This result also shows that food kind is not more important than the factors like service, taste, and surprise for FDR customers. Moreover, I determined that the most positive sentiments were directed to service by the customers while the most negative sentiments were directed to food kind. Furthermore, with the findings of this study, the collection of the most negative sentiments in food kind indicates that this feature is a confusing element for some customers in FDRs. In this respect, this study supports the claim expressed by Knutson et al. (1996) that a restaurant has little room to exceed expectations but has too much room to fail.

In the restaurant sample, I found that the highest performance was achieved by SVM in bigram, one of the n-gram language models. This result verifies the debate (Pang & Lee, 2008) whether high-order n-grams are useful properties. In the restaurant sample, Luo and Xu (2019) found that the performance of SVM + Fuzzy Domain

Ontology algorithms is higher together. Akhtar et al. (2017) found that the Conditional Random Field algorithm provided the highest performance followed by SVM. Kang et al. (2012) found that the unigram + bigram feature for sentiment analysis can be effective for the sentiment dictionary performance including unigrams and bigrams in the sample of restaurant reviews, and that the improved Naïve Bayes algorithm they suggest achieves better than the SVM algorithm.

To the best of my knowledge, the n-gram language models (unigram, bigram, trigram, and quadrigram) were used for the first time in this study for the best model in a restaurant sample, and Cohen's kappa was used in addition to other parameters (recall, precision, F-measure, and accuracy) in performance measurement. In this perspective, this paper combines the fields of hospitality and informatics by examining the issue from a technical perspective, as well as the usefulness of online restaurant reviews for the hospitality industry.

This paper provides some practical contributions to FDRs to understand their customers' sentiments and thoughts. First, ABSA can be used as a tool to assess the business and its products or services from the customers' perspective. This tool can also be a source of feedback for understanding the customer's expectations from the business and detecting the good and bad practices applied by the business and its competitors. In addition, the topic models determined in this study regarding the perception of service quality for FDRs constitute the communication process between businesses and customers. Moreover, monitoring of online portals customer reviews are posted and reputation management to be carried out through these portals will facilitate the communication process. Furthermore, price was not over emphasized among the topic models obtained in this study, so I can state that price is not one of the priorities of customers in FDRs. In other words, customers are willing to pay a high price by focusing on experience in FDRs. In this perspective, it will be a useful management approach for FDRs to present services beyond customers' expectations leading to unforgettable moments and the psychological quality suggested by this study.

Limitations and Future Studies

This study has some limitations. First, the research is limited to 25 restaurants ranked in TripAdvisor's Best Fine Dining Restaurants-World. The study can be repeated by using different online platforms in the sample of FDRs or other concepts related to food and beverage. Second, although TripAdvisor offers reviewers to enter this information, it is not a requirement. Due to the lack of this information for the entire dataset, I could not compare the reviews in terms of demographic variables. Third, the density of the dataset makes it difficult to analyze the topic models for the perception of service quality in detail. For the detailed analysis of topic models, each topic model can be a separate research topic in future studies. Fourth, in the performance test for sentiment polarity, the fact that the positive labeled data was more than the negative labeled data made it difficult to predict the negative labeled data for the controlled machine algorithms. In order to overcome this problem, the reviews of these two sentiments can be examined in close proportions in future studies.

Declaration of Interests

None.

REFERENCES

Akhtar, M. S., Gupta, D., Ekbal, A., & Bhattacharyya, P. (2017). Feature selection and ensemble construction: A two-step method for aspect based sentiment analysis. *Knowledge-Based Systems*, 125, 116–135. https://doi.org/10.1016/j.knosys.2017.03.020

Aktas-Polat, S., & Polat, S. (2022). Discovery of factors affecting tourists' fine dining experiences at five-star hotel restaurants in Istanbul. *British Food Journal*, 124(1), 221-238. https://doi.org/10.1108/BFJ-02-2021-0138

Aktas-Polat, S., & Polat, S. (2020). A theoretical analysis of food meaning in anthropology and sociology. *Tourism: An International Interdisciplinary Journal*, 68(3), 278–293. https://doi.org/10.37741/t.68.3.3

Alaei, A. R., Becken, S., & Stantic B. (2019). Sentiment analysis in tourism: Capitalizing on big data. *Journal of Travel Research*, 58(2), 175–191. https://doi.org/10.1177/0047287517747753

Antun, J. M., Frash, R. E., Costen, W., & Runyan, R. C. (2010). Accurately assessing expectations most important to restaurant patrons: The creation of the DinEX Scale. *Journal of Foodservice Business Research*, 13(4), 360–379. https://doi.org/10.1080/15378020.2010.524539

Bagheri, A., Saraee, M., & De Jong, F. (2014). ADM-LDA: An aspect detection model based on topic modelling using the structure of review sentences. *Journal of Information Science*, 40(5), 621–636.

Bengio, Y., & Grandvalet, Y. (2004). No unbiased estimator of the variance of k-fold cross-validation. *Journal of Machine Learning Research*, 5, 1089–1105.

Bermingham, L., & Lee, I. (2014). Spatio-temporal sequential pattern mining for tourism sciences. *Procedia Computer Science*, 29, 379–389.

Berry, L. L., Parasuraman, A., & Zeithaml, V. A. (1988). The service-quality puzzle. *Business Horizons*, 31(5), 35–43.

Bhuyan, M. P., & Sarma, S. K. (2019). An n-gram based model for predicting of word-formation in Assamese language. *Journal of Information and Optimization Sciences*, 40(2), 427–440, https://doi.org/10.1080/02522667.2019.1580883

Bitner, M. J. (1992). Servicescapes: The impact of physical surroundings on customers and employees. *Journal of Marketing*, 56(2), 57–71. https://doi.org/10.2307/1252042

Bojanic, D. C., & Rosen, L. D. (1994). Measuring service quality in restaurants: An application of the SERVQUAL instrument. *Journal of Hospitality & Tourism Research*, 18(1), 3–14. https://doi.org/10.1177/109634809401800102

Chase R. B., & Apte U. M. (2007). A history of research in service operations: What's the big idea? *Journal of Operations Management*, 25(2), 375–386.

Cheng, C. C., Chen, C. T., Hsu, F. S., & Hu, H. Y. (2012). Enhancing service quality improvement strategies of fine-dining restaurants: New insights from integrating a two-phase decision-making model of IPGA and DEMATEL analysis. *International Journal of Hospitality Management*, 31(4), 1155–1166.

Cohen, J. (1960). A coefficient of agreement for nominal scales. *Educational and Psychological Measurement*, 20(1), 37–46.

Cronin, J. J., & Taylor, S. A. (1992). Measuring service quality: A reexamination and extension. *Journal of Marketing*, 56(3), 55–68.

Dey, A., Jenamani, M., & Thakkar, J. J. (2018). Senti-n-gram : An n-gram lexicon for sentiment analysis. *Expert Systems with Applications*, 103, 92–105. https://doi.org/ 10.1016/j.eswa.2018.03.004

Duan, W., Yu, Y., Cao, Q., & Levy, S. (2016). Exploring the impact of social media on hotel service performance. *Cornell Hospitality Quarterly*, 57(3), 282–296. https://doi.org/10.1177/1938965515620483

Feldman, R. (2013). Techniques and applications for sentiment analysis. *Communications of the ACM*, 56(4), 82–89. https://doi.org/10.1145/2436256.2436274

Fu, Y., Hao, J. X., Li, X., & Hsu. C. H. (2019). Predictive accuracy of sentiment analytics for tourism: A meta learning perspective on Chinese travel news. *Journal of Travel Research*, 58(4), 666–679. https://doi.org/10.1177/0047287518772361

Gagić, S., Tešanović, D., & Jovičić, A. (2013). The vital components of restaurant quality that affect guest satisfaction. *Turizam*, 17(4), 166–176.

García-Pablos, A., Cuadros, M., & Rigau, G. (2018). W2VLDA: Almost unsupervised system for aspect based sentiment analysis. *Expert Systems with Applications*, 91, 127–137. https://doi.org/10.1016/j.eswa.2017.08.049

Geler, Z., Savić, M., Bratić, B., Kurbalija, V., Ivanović, M., & Dai, W. (2021). Sentiment prediction based on analysis of customers assessments in food serving businesses. *Connection Science*, 33(3), 674–692. https://doi.org/10.1080/09540091.2020.1870436

Grönroos, C. (1984). A service quality model and its market implications. *European Journal of Marketing*, 18(4), 36–44.

Guo, Y., Barnes, S. J., & Jia, Q. (2017). Mining meaning from online ratings and reviews: Tourist satisfaction analysis using latent Dirichlet allocation. *Tourism Management*, 59, 467–483. http://dx.doi.org/10.1016/j.tourman.2016.09.009

Hansen, K. V. (2014). Development of SERVQUAL and DINESERV for measuring meal experiences in eating establishments. *Scandinavian Journal of Hospitality and Tourism*, 14(2), 116–134. https://doi.org/10.1080/15022250.2014.886094

Hearst, M. (2003). What is text mining. SIMS, UC Berkeley.

Hu, M., & Liu, B. (2004). Mining and summarizing customer reviews. In Proceedings of the tenth ACM SIGKDD international conference on Knowledge discovery and data mining. *Association for Computing Machinery*, pp. 168–177.

Huang, C. H. (2003). A comparison of the service quality perspectives between restaurant customers and restaurateurs in Taiwan. *Journal of Foodservice Business Research*, 6(2), 87–104. https://doi.org/10.1300/J369v06n02_07

Hwang, J., & Ok, C. (2013). The antecedents and consequence of consumer attitudes toward restaurant brands: A comparative study between casual and fine dining restaurants. *International Journal of Hospitality Management*, 32, 121–131. https://doi.org/10.1016/j.ijhm.2012.05.002

James, T. L., Calderon, E. D. V., & Cook, D. F. (2017). Exploring patient perceptions of healthcare service quality through analysis of unstructured feedback. *Expert Systems with Applications*, 71, 479–492. https://doi.org/10.1016/j.eswa.2016.11.004

Jin, N. P., Line, N. D., & Goh, B. (2013). Experiential value, relationship quality, and customer loyalty in full-service restaurants: The moderating role of gender. *Journal of Hospitality Marketing & Management*, 22(7), 679–700. http://dx.doi.org/10.1080/19368623.2013.723799

Johns, N., & Tyas, P. (1996). Use of service-quality gap theory to differentiate between foodservice outlets. *The Service Industries Journal*, 16(3), 321–346.

Johns, N., & Howard, A. (1998). Customer expectations versus perceptions of service performance in the foodservice industry. *International Journal of Service Industry Management*, 9(3), 248–265. https://doi.org/10.1108/09564239810223556

Kala, D. (2020). Examining the impact of food attributes and restaurant services on tourist satisfaction: Evidence from mountainous state of India. *Journal of Quality Assurance in Hospitality & Tourism*, 21(4), 430–453. https://doi.org/10.1080/1528008X.2019.1672235

Kang, H., Yoo, S. J., & Han, D. (2012). Senti-lexicon and improved Naïve Bayes algorithms for sentiment analysis of restaurant reviews. *Expert Systems with Applications*, 39(5), 6000–6010. https://doi.org/10.1016/j.eswa.2011.11.107

Keith, N. K., & Simmers, C. S. (2011). Measuring service quality perceptions of restaurant experiences: The disparity between comment cards and DINESERV. *Journal of Foodservice Business Research*, 14(1), 20–32. https://doi.org/10.1080/15378020.2011.548209

Kim, H. J., McCahon, C., & Miller, J. (2003). Assessing service quality in Korean casual-dining restaurants using DINESERV. *Journal of Foodservice Business Research*, 6(1), 67–86. https://doi.org/10.1300/j369v06n01_05

Kim, K., Park, O. J., Yun, S., & Yun, H. (2017). What makes tourists feel negatively about tourism destinations? Application of hybrid text mining methodology to smart destination management. *Technological Forecasting and Social Change*, 123, 362–369. https://doi.org/10.1016/j.techfore.2017.01.001

Kirilenko, A. P., Stepchenkova, S. O., Kim, H., & Li, X. (2018). Automated sentiment analysis in tourism: Comparison of approaches. *Journal of Travel Research*, 57(8), 1012–1025. https://doi.org/10.1177/2F0047287517729757

Knutson, B. J., Stevens, P., & Patton, M. (1996). DINESERV: Measuring service quality in quick service, casual/theme, and fine dining restaurants. *Journal of Hospitality & Leisure Marketing*, 3(2), 35–44. https://doi.org/10.1300/J150v03n02_04

Korfiatis, N., Stamolampros, P., Kourouthanassis, P., & Sagiadinos, V. (2018). Measuring service quality from unstructured data: A topic modeling application on airline passengers' online reviews. *Expert Systems with Applications*, 116, 472–486. https://doi.org/10.1016/j.eswa.2018.09.037

Kwon, W., Lee, M., & Back, K. J. (2020). Exploring the underlying factors of customer value in restaurants: A machine learning approach. *International Journal of Hospitality Management*, 91, 102643. https://doi.org/10.1016/j.ijhm.2020.102643

Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics,* 33(1), 159–174.

Lee, L., Lee, M. J., & Dewald, B. (2016). Measuring the customers' perception of tangible service quality in the restaurant industry: An emphasis on the upscale dining segment. *Journal of Foodservice Business Research*, 19(1), 21–38. http://dx.doi.org/10.1080/15378020.2016.1129216

Li, H., Liu, H., & Zhang, Z. (2020). Online persuasion of review emotional intensity: A text mining analysis of restaurant reviews. *International Journal of Hospitality Management*, 89, 102558. https://doi.org/10.1016/j.ijhm.2020.102558

Lim, J., & Lee, H. C. (2020). Comparisons of service quality perceptions between full service carriers and low cost carriers in airline travel. *Current Issues in Tourism*, 23(10), 1261–1276. https://doi.org/10.1080/13683500.2019.1604638

Lisnawati, L., & Astawa, I. P. (2020). Service quality analisys of 'Table8' Chinese restaurant at the Mulia, Mulia resort and Mulia villas. *Journal of Applied Sciences in Travel and Hospitality*, 3(1), 22–29.

Luo, Y., He, J., Mou, Y., Wang, J., & Liu, T. (2021). Exploring China's 5A global geoparks through online tourism reviews: A mining model based on machine learning approach. *Tourism Management Perspectives*, 37, 100769. https://doi.org/10.1016/j.tmp.2020.100769

Luo, Y., & Xu, X. (2019). Predicting the helpfulness of online restaurant reviews using different machine learning algorithms: A case study of Yelp. *Sustainability*, 11(19), 5254. http://dx.doi.org/10.3390/su11195254

MacLaurin, D. J., & MacLaurin, T. L. (2000). Customer perceptions of Singapore's theme restaurants. *Cornell Hotel and Restaurant Administration Quarterly*, 41(3), 75–85. https://doi.org/10.1177/001088040004100319

Madanoglu, M. (2006). Validating restaurant service quality dimensions. *Journal of Foodservice Business Research*, 7(4), 127–147. https://doi.org/10.1300/J369v07n04_07

Marković, S., Raspor, S., & Šegarić, K. (2010). Does restaurant performance meet customers' ecpectations? An assessment of restaurant service quality using a modified dineserv approach. *Tourism and Hospitality Management*, 16(2), 181–195.

McHugh, M. L. (2012). Interrater reliability: The kappa statistic. *Biochemia Medica*, 22(3), 276–282.

Mejia, J., Mankad, S., & Gopal, A. (2020). Service quality using text mining: Measurement and consequences. *Manufacturing & Service Operations Management*, 23(6), 1354–1372 https://doi.org/10.1287/msom.2020.0883

Moro, S., Lopes, R. J., Esmerado, J., & Botelho, M. (2020). Service quality in airport hotel chains through the lens of online reviewers. *Journal of Retailing and Consumer Services*, 56, 102193. https://doi.org/10.1016/j.jretconser.2020.102193

Palese, B., & Piccoli, G. (2016). Online reviews as a measure of service quality. In 2016 *Pre-ICIS SIGDSA/IFIP WG8.* 3 Symposium, Dublin 2016.

Palese, B., & Usai, A. (2018). The relative importance of service quality dimensions in e-commerce experiences. *International Journal of Information Management*, 40, 132–140. https://doi.org/10.1016/j.ijinfomgt.2018.02.001

Pang, B., Lee, L., & Vaithyanathan, S. (2002). Thumbs up? Sentiment classification using machine learning techniques. In Proceedings of the ACL-02 conference on empirical methods in natural language processing-Volume 10. *Association for Computational Linguistics*, pp. 79–86.

Pang, B., & Lee, L. (2004). A sentimental education: Sentiment analysis using subjectivity summarization based on minimum cuts. *In Proceedings of the 42nd annual meeting on Association for Computational Linguistics*. Association for Computational Linguistics.

Pang, B., & Lee, L. (2008). Opinion mining and sentiment analysis, *Foundations and Trends® in Information Retrieval*, 2(1–2), 1–135. https://doi.org/10.1561/1500000001s

Pantelidis, I. S. (2010). Electronic meal experience: A content analysis of online restaurant comments. *Cornell Hospitality Quarterly*, 51(4), 483–491. https://doi.org/10.1177/1938965510378574

Parasuraman, A., Zeithaml, V. A., & Berry, L. L. (1985). A conceptual model of service quality and its ımplications for future research. *The Journal of Marketing*, 49(4), 41–50.

Parasuraman, A., Zeithaml, V. A., & Berry, L. L. (1988). Servqual: A multiple-item scale for measuring consumer perceptions of service quality. *Journal of Retailing*, 64(1), 12–40.

Philander, K., & Zhong, Y. Y. (2016). Twitter sentiment analysis: Capturing sentiment from integrated resort tweets. International *Journal of Hospitality Management*, 55, 16–24. http://dx.doi.org/10.1016/j.ijhm.2016.02.001

Quinlan, J. R. (1993). C4.5: *Programs for machine learning.* Morgan Kaufmann Publishers.

Raajpoot, N. A. (2002). TANGSERV: A multi-item scale for measuring customer tangible quality in foodservice industry. *Journal of Foodservice Business Research*, 5(2), 109–127. https://doi.org/10.1300/J369v05n02_08

Rust, R. T., & Oliver, R. L. (1994). *Service quality: Insights and managerial implications from the frontier*. R. T. Rust, & R. L. Oliver (Ed.s.), Service quality: New directions in theory and practice. Sage Publications.

Ryu, K., & Jang, S. S. (2008). DINESCAPE: A scale for customers' perception of dining environments. *Journal of Foodservice Business Research*, 11(1), 2–22. https://doi.org/10.1080/15378020801926551

Schuckert, M., Liu, X., & Law, R. (2015). Hospitality and tourism online reviews: Recent trends and future directions. *Journal of Travel & Tourism Marketing*, 32(5), 608–621. https://doi.org/10.1080/10548408.2014.933154

Stevens, P., Knutson, B., & Patton, M. (1995). DINESERV: A tool for measuring service quality in restaurants. *The Cornell Hotel and Restaurant Administration Quarterly*, 36(2), 5–60. https://doi.org/10.1016/0010-8804(95)93844-K

Sutherland, I., Sim, Y., Lee, S. K., Byun, J., & Kiatkawsin, K. (2020). Topic modeling of online accommodation reviews via latent Dirichlet allocation. *Sustainability*, 12, 1821. https://doi.org/10.3390/su12051821

Taecharungroj, V., & Mathayomchan, B. (2019). Analysing TripAdvisor reviews of tourist attractions in Phuket, Thailand. *Tourism Management*, 75, 550–568. https://doi.org/10.1016/j.tourman.2019.06.020

Taecharungroj, V., Warnaby, G., & Parker, C. (2021). Responding to the voice of the markets: an analysis of Tripadvisor reviews of UK retail markets. *Journal of Place Management and Development*, 14(2), 180-200, https://doi.org/10.1108/JPMD-02-2020-0016.

Tan, P. N., Steinbach, M., & Kumar, V. (2014). *Introduction to data mining*. Pearson Education Limited.

Tao, D., Yang, P., & Feng, H. (2020). Utilization of text mining as a big data analysis tool for food science and nutrition. *Comprehensive Reviews in Food Science and Food Safety*, 19(2), 875–894. https://doi.org/10.1111/1541-4337.12540

Tian, G., Lu, L., & McIntosh, C. (2021). What factors affect consumers’ dining sentiments and their ratings: Evidence from restaurant online review data. *Food Quality and Preference*, 88, 104060. https://doi.org/10.1016/j.foodqual.2020.104060

TripAdvisor (2021). Travelers’ choice best of the best. https://www.tripadvisor.com/TravelersChoice.

Wall, E. A., & Berry, L. L. (2007). The combined effects of the physical environment and employee behavior on customer perception of restaurant service quality. *Cornell Hotel and Restaurant Administration Quarterly*, 48(1), 59–69. https://doi.org/10.1177/0010880406297246

Yan, X., Wang, J., & Chau, M. (2015). Customer revisit intention to restaurants: Evidence from online reviews. *Information Systems Frontiers*, 17(3), 645–657. https://doi.org/10.1007/s10796-013-9446-5

Yousefpour, A., Ibrahim, R., Hamed, H. N. A., & Hajmohammadi, M. S. (2014). A comparative study in sentiment analysis. *Advances in Environmental Biology*, 8(13), 53–68.

Zeithaml, V. (1987). *Defining and relating price, perceived quality, and perceived value*. (Report No. 87–101). Marketing Science Institute.

Zhang, L., & Liu, B. (2014). *Aspect and entity extraction for opinion mining*. W. W. Chu (Ed.), Data mining and knowledge discovery for big data. Studies in big data, 1, 1–40. Springer. 10.1007/978-3-642-40837-3_1

Zhang, W., Xu, H., & Wan, W. (2012). Weakness finder: Find product weakness from Chinese reviews by using aspects based sentiment analysis. *Expert Systems with Applications*, 39(11), 10283–10291. https://doi.org/10.1016/j.eswa.2012.02.166

Zhao, G., Qian, X., Lei, X., & Mei, T. (2016). Service quality evaluation by exploring social users’ contextual information. *IEEE Transactions on Knowledge and Data Engineering*, 28(12), 3382–3394. https://doi.org/10.1109/TKDE.2016.2607172