PeerJ Computer Science

Determining the factors affecting customer satisfaction using an extraction-based feature selection approach

Weishen Wu[1] and Dalianus Riantama[2]

[1] Department of Information Management, Da-Yeh University, Changhua, Taiwan
[2] College of Management, Da-Yeh University, Changhua, Taiwan

ABSTRACT

The coronavirus disease 2019 (COVID-19) causes tremendous damages to the world, including threats to human's health and daily activities. Most industries have been affected by this pandemic, particularly the tourism industry. The online travel agencies (OTAs) have suffered from the global tourism market crisis by air travel lockdown in many countries. How online travel agencies can survive at stake and prepare for the post-COVID-19 future has emerged as an urgent issue. This study aims to examine the critical factors of customers' satisfaction to OTAs during the COVID-19 pandemic. A text mining method for feature selection, namely LASSO, was used to deal with online customer reviews and to extract factors that shape customers' satisfaction to OTAs. Results showed that refunds, promptness, easiness and assurance were ranked as the most competitive factors of customers' satisfaction, followed by bad reviews & cheap and excellent service & comparison. New factors to customers' satisfaction were revealed during the global tourism recession. Findings provide OTAs guidelines to reset services priorities during the pandemic crisis.

Subjects Data Mining and Machine Learning, Embedded Computing, Natural Language and Speech
Keywords COVID-19, Online travel agencies, Text mining, LASSO, Feature selection, Customer satisfaction

Submitted 24 June 2021
Accepted 17 December 2021
Published 25 January 2022

Corresponding author
Dalianus Riantama,
dalianusriantama@gmail.com

Academic editor
Vimal Shanmuganathan

Additional Information and Declarations can be found on page 12

DOI **10.7717/peerj-cs.850**

INTRODUCTION

Online travel agencies (OTAs) are online business that facilitates customers to purchase travel, hospitality, and tourism products/services from providers (*e.g.*, airlines, hotels, rental cars, restaurants, cruises, package holidays, etc.) and gets commission charged on transactions as an intermediary (*Long & Shi, 2017*). During the coronavirus disease 2019 (COVID-19) pandemic worldwide in 2020, OTAs have been hit hard (*Sigala, 2020*). Given the negative impacts on the tourism industry in such a crisis, customer satisfaction turns out to be crucial thus real-time research on it is desperately demanded (*Sharma, Sharma & Chaudhary, 2020*; *Zhang, Hou & Li, 2020*). Several OTAs started to cooperate with tourism suppliers to promote customer satisfaction during the COVID-19 pandemic (*Hao, Xiao & Chon, 2020*).

Most of the studies adopted questionnaires to obtain customers' voices to measure the factors that influence customer satisfaction in various tourism contexts, such as OTAs (*Hao et al., 2015*; *Rajaobelina, 2018*), hotels (*Davras & Caber, 2019*; *Nunkoo et al., 2020*), airlines

(*Tahanisaz & Shokuhyar, 2020*) and restaurants (*Gopi & Samat, 2020*). For the survey method, respondents may not pay attention to every item or randomly answer questions resulting in incomplete data (*Evans & Mathur, 2018*). This study uses online customer reviews (OCRs) that can lower the inaccuracy of artificial responses given by customers to questionnaire surveys (*Sánchez-Franco, Navarro-García & Rondán-Cataluña, 2019*). OCRs are the user-generated content containing text comments and rating scores of companies or brands that are posted on e-vendor websites or third-party websites (*Mudambi & Schuff, 2010*). Few researchers utilized OCRs in the domain of online travel agencies but not to understand customer satisfaction (*Hou et al., 2019*). In addition, the questionnaire survey demands researchers to identify the satisfaction's factors in advance whereas a gap between what researchers and managers believe is important and what customers say is important in the evaluation and selection of services (*Lockyer, 2005*). In contrast to previous studies, instead of identifying customer satisfaction's factors in advance, this research goes differently to search for customer satisfaction's factors blindfold. This study uses OCRs to apply exploratory research to understand customer satisfaction's factors, and the outcomes can be more reliable. Exploratory research is suitable for this study because we believe that we could not use our past knowledge to judge for specific situations such as COVID-19 circumstances.

Drawing from the literature on impression formation, it is important that researchers further investigate the first question of what causes customer satisfaction among OTAs' customers. To the OTAs, customer satisfaction is an antecedent to customer loyalty, repurchasing, and positive/negative OCRs (*Rianthong, Dumrongsiri & Kohda, 2016*; *Long & Shi, 2017*; *Cui, Lin & Qu, 2018*; *Brun et al., 2020*; *Sharma, Sharma & Chaudhary, 2020*). The second question of what is from the most to the least important ranked of customer satisfaction's attributes. As a result, OTAs can better set priorities for the attributes that are most important to customers while also improving cost performance. By answering those questions, this study contributes to the literature in two respects. This study aims to reveal and rank the significant factors of OTA customer satisfaction during the COVID-19 outbreak.

The remainder of this article is organized as follows. This study first presents an outline of the foundation of this examination and relevant literature. This study extracts OCRs and adopts a text mining approach to deal with them. Next, this research investigates customer satisfaction's factors using a multimethod approach applying big data sets from the largest OTA in the world. Finally, this study discusses the key findings and practical implications for OTAs and considers future examination necessities.

LITERATURE REVIEWS

Customer satisfaction toward online travel agencies

The concept of customer satisfaction covers the expectation/disconfirmation paradigm, the norm view, the equity view, and the perceived overall performance (*Yoon & Uysal, 2005*). The theoretical foundation of this research is based on expectation/disconfirmation theory. There are two scenarios for expectation/disconfirmation theory: affirmation (satisfaction)

if the perceived outcome meets expectations; and negative disconfirmation (dissatisfaction) if expectations are not reached (*Yüksel & Yüksel, 2001*). Previous studies show that factors influencing customers' pre-purchasing expectations consist of product- and service-related factors and customer-related factors. In the context of OTA, the product- and service-related factors include website reputation, available choices, and product price (*Chang, Hsu & Lan, 2019*) and influence customer expectations (*Ha & Janda, 2016*; *Kim et al., 2020*).

Service quality attributes were the most factors analyzed by previous studies to understand customer satisfaction as shown in Table 1, largely ignoring external factors. It is unknown whether external factors have an impact on customer satisfaction toward OTAs. Nowadays, the hospitality and tourism industry is very influenced by the rapid development of information technology. The internet makes external factors such as online customer reviews hold a big portion to affect customer satisfaction (*Sharma, Sharma & Chaudhary, 2020*; *Wang et al., 2020*).

Online customer reviews

OCRs provide a rich source of data to extract the dimensions of customer satisfaction for tourism sectors (*Chen et al., 2019*; *Hlee et al., 2020*; *Joung, Kim & Kim, 2021*; *Lien, Wen & Wu, 2011*; *Zinko et al., 2021*). The results of the studies using OCRs ought to be more dependable and exact than those statistical results acquired from conventional satisfaction surveys dependent on little data samples (*Sánchez-Franco, Navarro-García & Rondán-Cataluña, 2019*). In addition, when the social distancing was carried out in the pandemic, readers' perceptions toward certain products or services mainly relied on OCRs (*Hernández-Ortega, 2018*).

OCRs usually contain text comments and overall ratings. These comments demonstrate customer satisfaction's attributes, and the overall ratings show customers' overall satisfaction (*Xu, 2020*). *Tao & Kim (2019)* used OCRs to find a new attribute of customer satisfaction which is onshore cruiser experiences attributes. *Situmeang, de Boer & Zhang (2020)* comprehended customer satisfaction using OCRs and affirmed OCRs can develop a sustainable strategy for the restaurant industry. Based on the above findings, this study utilizes OCRs to discover the vital attributes of OTA customer satisfaction.

A Text mining approach for pre-processing

Text mining is a knowledge exploration approach that consolidates techniques of natural language processing, information retrieval, machine learning, and data mining (*Yang et al., 2018*; *Zhou & Xue, 2020*). The essential task of text mining is to transform texts into numerical data for analysis through natural language processing including editing, analyzing, and organizing an enormous number of texts to provide explicit information (*Sullivan, 2001*). Previous studies found that text mining was an efficient way to obtain key issues from an enormous number of OCRs and customers' thoughts can be demonstrated all the more plainly (*Xu & Li, 2016*; *Chiu & Lin, 2018*). Compared with manual content analysis, text mining has relevant advantages such as less time and human works to perform analysis (*Guo et al., 2016*) and extraction of new variables (*Hong & Park, 2019*).

Text mining techniques have been applied in different subjects particularly in tourism and hospitality research. *Jia (2018)* proposed a pre-processing process to analyze restaurant

Table 1 Review of literature on satisfaction in tourism sectors.

References	Context	Factors/Predictors/Antecedents	Findings
Kim & Lee (2004)	OTA	Structure & ease of use, information content, usefulness & reputation, and security	Information content was found to be the most important factor in explaining customer satisfaction.
Chen & Kao (2010)	OTA	Process quality and outcome quality	Process quality and outcome quality influence customer satisfaction.
Tsang, Lai & Law (2010)	OTA	Customer relationship, safety & security, website functionality, fulfillment & responsiveness, information quality & content, appearance, and presentation	Website functionality, information quality & content, safety & security, and customer relationship influence customer satisfaction.
Hsu, Chang & Chen (2012)	OTA	Perceived flow and perceived playfulness	Perceived flow and customers' perceived playfulness affect satisfaction.
Ting, Chen & Lee (2013)	OTA	technology acceptance, perceived risk, reduced transaction cost, and service quality	Customers' e-satisfaction is influenced by service quality and online risk.
Pereira, de Fátima Salgueiro & Rita (2017)	OTA	Online routine, customers' innovativeness, website's image perceptions, and online knowledge	Routine, website's image, and knowledge significantly affect e-customer satisfaction.
Ju et al. (2019)	Airbnb	n/a	The facility produces distinctive, website, and host effects on customer satisfaction.
Sthapit et al. (2020)	Airbnb	Consumption's values (functional, social, and emotional), co-creation, and information overload	The absence of information overload and co-creation contribute to satisfaction with using the Airbnb website.
Prassida, Hsu & Chang (2021)	OTA and Hotel	Service quality and the perceived value	The perceived value of offline services and online service quality are crucial influence customer satisfaction.

customers' reviews and present insights into the analysis of reviews. *Cheng & Jin (2019)* identified 'price' as a key influencer to Airbnb with a text mining approach on OCRs. This study employs a text mining approach to transform OCRs into numerical data prepared for the feature selection process.

Feature selection with least absolute shrinkage and selection operator (LASSO)

Feature selection is a process of looking for the best subset of characteristics, from the original set according to the given goal of processing and criteria (*Swiniarski & Skowron, 2003*). Feature selection has two purposes which are to avoid the curse of dimensionality in modeling and to get important features. Its process is to eliminate unimportant features that can decrease the difficulty of learning tasks (*Kwok, Zhou & Xu, 2015*). Due to the frequent long length, generous number, and open structure of online textual reviews, extracting key points from textual reviews can be challenging and complex (*Gandomi &*

Haider, 2015). The questions are which features are to be included in the model, and which feature selection algorithms can be employed. The existing solutions of feature selection can be separated into the filter, wrapper, and embedded methods. The filter method is a pre-processing stage and uses criteria not involving any learning machine and, by doing that, it does not consider the impacts of a chosen feature subset (*Kohavi & John, 1998*; *Guyon & Elisseefl, 2006*; *Lal et al., 2006*). The wrapper method assesses a subset of features according to the accuracy of a given predictor (*Kohavi & John, 1998*; *Guyon & Elisseeff, 2003*). The embedded methods of feature selection are suitable for the process of training and to give learning machines (*Guyon & Elisseeff, 2003*). Filter and wrapper methods do not evaluate the feature sets iteratively, in contrast, the embedded method is more robust in over-fitting data (*Cai et al., 2018*). One typical algorithm of the embedded methods is called the least absolute shrinkage and selection operator (LASSO) (*Tibshirani, 1996*).

LASSO is a regression method that involves setting the absolute size of the regression coefficients which does regression and feature selection simultaneously to enhance interpretability of the statistical model it produces (*Tibshirani, 1996*). LASSO forces a limit on the sum of absolute values of the regression coefficients, enabling some coefficients to be zero, exposing unimportant features, so those coefficients of important features are not zero. The principal feature of LASSO is that the pressure factor and the feature selection can be automatically cultivated in the evaluation process (*Huang, Wang & Kochenberger, 2017*). Through a variable selection procedure with shrinkage of regression coefficients to zero then picking the most fitted coefficients in the linear regression, LASSO controls the model complexity and increases the selection performance (*Sant'Anna, Caldeira & Filomena, 2020*). Past research confirmed a better result can be accomplished by using LASSO.

Previous research has shown that LASSO outperforms other algorithms in terms of results. *Chang et al. (2019)* used support vector machines (SVM) and back-propagation neural networks (BPN) to compare LASSO and decision tree (DT) in order to find the most critical un-revisit intention factors, and found that LASSO had higher accuracy than DT. *Dastjerdi, Foroghi & Kiani (2019)* predicted a manager's fraud risk and came up with a LASSO result that was much more precise than the Convex Optimization (CVX). After being analyzed by support vector machines (SVM), *Chang et al. (2020)* discovered that LASSO obtained superior accuracy compared to support vector machines recursive feature elimination (SVM-RFE) in order to determine the most important factors toward customers' trust in O2O models. This study employs LASSO to do feature selection due to its powerful algorithm which enables to get the most important variables of the OTA customer satisfaction from OCRs.

RESEARCH METHODS

In line with previous studies (*Chang et al., 2020*; *Chen et al., 2021*), the feature selection consists of the following five steps; (1) data collection, (2) data pre-processing, (3) generate TF-IDF, (4) Lasso, and (5) words labeling (Fig. 1). Details of the process are described as follows.

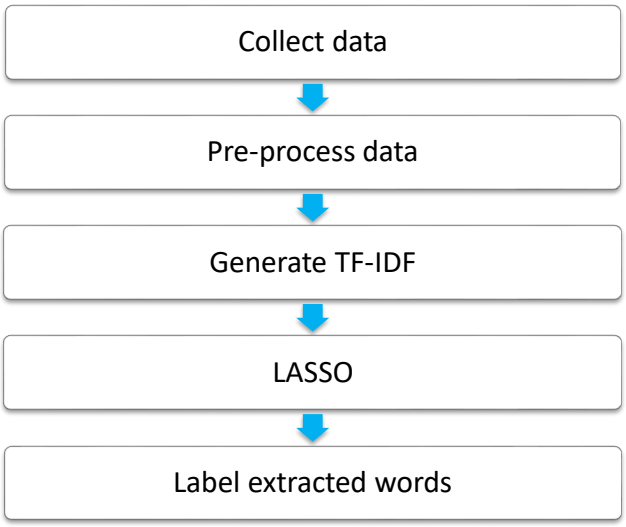

Figure 1 **Steps of feature selection.**
Full-size DOI: 10.7717/peerjcs.850/fig-1

Data collection

This study collected and analyzed customers' opinions toward a well-known OTA because it operates worldwide (*Trefis Team, Great Speculations, 2019*). This OTA was available in 43 languages and offered 28 million accommodations at 15 thousand destinations in 226 countries and territories in December 2020. OCRs were considered as a source of data because they reflected alternative customers' perceptions. Trustpilot.com (https://www.trustpilot.com/) is an open online platform for evaluating services, companies, or brands by customers. Trustpilot.com provided more OCRs than other similar websites during the data collection period. *Singh & Söderlund (2020)* also collected retailers' OCRs in the UK from Trustpilot.com. Therefore, the OTA customers' reviews on Trustpilot.com posted in English were chosen as samples in this study. To ensure these reviews represent the majority of customers' voices, *Chiu & Lin (2018)* suggested that the minimum reviews have better results with more than 50 samples. A total of 1,313 OCRs with comment texts and overall ratings (1–5 scores) were obtained, from March to August in 2020 during the COVID-19 outbreak.

Data pre-processing

Online customer reviews commonly appear with long sentences. In order to get fewer words but probably more important words. The TF-IDF process was applied to clean the sentences into pieces by pieces of words based on their occurrences. Along the process, the words with low occurrence would be removed. The data pre-processing was performed by the data analytics software, namely RapidMiner Studio® 9.4r. A tokenization function was applied to remove unrelated characters, symbols, emoticons, and stop words, such as "the", "are", "that", etc., and to reorganize the texts into lowercase letters. This function was also used to avoid words less than three letters that could not provide enough significant information, such as "on", "at", "no", etc. The texts were tokenized with non-letter

separators that separated the comments into small pieces. Further, a stem method was applied to the root of the token, for example, "simplistic" and "simplicity" were purified into the single token "simple" resulting in a single meaning of words. Segment corpus with bigram in which two words were often found together throughout the document, such as "full_refund", "excellent_service". Then a pruning method was applied by which any words appearing less than five times in the dataset were removed because these words were mentioned less which meant having a less significant contribution to the model. Finally, the term frequency-inverse document frequency (TF-IDF), the relative frequency of a certain word in a specific document (*Ramos, 2003*; *Sezgen, Mason & Mayer, 2019*), was ready to be analyzed. TF-IDF was confirmed to be an effective method for word weighting in information retrieval (*Sebastiani, 2002*). *Sezgen, Mason & Mayer (2019)* applied TF-IDF to deal with online customer reviews to analyze the reviews further.TF-IDF is defined as follows.

$$idf_i = \log 2(\frac{N}{n_i}) + 1. \tag{1}$$

TF-IDF (weighted) score is calculated by;

$$w_{ij} = tf_{ij} \times idf_i. \tag{2}$$

In Eq. (1), N = the number of total documents and n_i = the term frequency of term i in the overall documents. In Eq. (2), tf_{ij} refers to the number of occurrences of term i in document j and idf_i represents the general significance of term i in the overall documents. TF-IDF is a metric that multiplies the two quantities tf and idf. This method was applied to weight which words were most frequently shown in one single review. When a word's TF-IDF score is higher, it demonstrates the word appears frequently in the part of documents (*Chen et al., 2016*; *Sebastiani, 2002*). The most frequent words would be analyzed further. In this study, TF-IDF was used to calculate the weights of words in a document. Finally, a TF-IDF weighted with 5,409 (selected words) × 1,313 (data samples) term-by-document matrix was established. The TF-IDF result was used by LASSO for selecting the important words.

Trustpilot.com allows customers to give overall ratings from 1 to 5 scores for a subject. *Farhadloo, Patterson & Rolland (2016)* transformed the ratings into a binary scale, with an overall rating score of 1, 2, or 3 being marked 0 and scores of 4 or 5 being marked 1. This method converts the 5-point scale into a 2-point binary scale representing bad *versus* good satisfaction (0 = unsatisfied and 1 = satisfied), and its robustness was confirmed by previous studies (*Atalık, Bakır & Akan, 2019*; *Tao & Kim, 2019*). The dependent variable in the LASSO method adopted the binary mode (zero and one) which is more precise and powerful than the continuous-dependent mode (*Dastjerdi, Foroghi & Kiani, 2019*). In this study, a binary method was used to mirror customer satisfaction's scores.

LASSO

Once the TF-IDF was established, LASSO was run by Matlab® software. It performed regression and feature selection functions simultaneously to extract the significant features

considering the following selection criteria, as shown in Eq. (3), where x is the explanatory variable, T is the number of data and λ is the adjustment coefficient.

$$\min = \sum_{t=1}^{T}(y_t - \beta_0 - \beta_1 x_{1,t} - \ldots - \beta_k x_{k,t})^2, \text{ s.t. } \sum_{j=1}^{k}\left|\beta_j\right| \leq \lambda. \quad (3)$$

According to Eq. (3), a regression parameter value namely βi is limited by a specific penalty selection benchmark, and afterward, the suitable variables are chosen. Given a k-explained transformation, the parameter estimate '\hat{\beta}' is influenced by the value of λ. When the λ's value approaches infinity, the estimate of the parameter \hat{\beta} is not limited, and the estimate is the value determined by the least-squares method. The contrary situation is when the λ is adjusted to 0, all parameter estimates become 0. The explanatory variable x, which is closely connected with y_t, would vary and differ from zero as the value of λ is gradually increased from zero, suggesting that the explanatory variable is significant. As a result, in this experiment, the premise for finding essential features is to see if the coefficient is 0, and if the coefficient is not 0, the feature is considered significant.

Identifying factors of customer satisfaction

After gathering the relevant keywords with LASSO, the following step was to classify them using a five-fold cross-validation experiment to establish the important factors based on their frequency of occurrence. The essential idea of the five-fold cross-validation experiment is that the sample data set is randomly partitioned into five mutually exclusive subsets (the folds). The technique was carried out in stages, with one subset serving as a testing subset and the other four serving as training subsets, and it ran in turn. While the group experiment approach was not relevant during the procedure, the five-fold cross-validation experiment ensured that every measurement was used for the objectives of training, testing, and validating. The five-fold cross-validation experiment was used to rank the important words based on their occurrence frequencies. When a word appears more times the more significant the word is *Lim & Kim (2020)*. *Chang et al. (2019)* and *Chang et al. (2020)* applied a five-fold cross-validation experiment to rank selected features.

EXPERIMENTAL RESULTS

LASSO results

In the parameter setting of LASSO, built-in functions in Matlab® were employed to filter out the essential words. would impose some words' regression coefficients to zero which means these words are not relevant to the regression model (*Zhao & Yu, 2006*; *Makarov et al., 2019*; *Wang, 2021*). Simply put, the words with regression coefficients zero were considered as not important words to influence customer satisfaction. Whereas, words with regression coefficients that are not zero can be considered as important words to influence customer satisfaction (*Zhang & Huang, 2008*). Since the five-fold cross-validation experiment approach was applied, the dataset was split into five equivalent parts. The five parts were run each by parameter setting of LASSO. With a five-fold cross-validation experiment approach, the results were also obtained five results as shown in Table 2 which is Fold#1, Fold#2, so on.

Table 2 LASSO Results.

Extracted keywords	Fold#1	Fold#2	Fold#3	Fold#4	Fold#5	Frequency
thank_you	2.5884113	2.5038914	2.4260114	2.5038914	2.5038914	5
never_had	2.3115838	2.0744914	1.9410475	2.0744914	2.0744914	5
great	1.8180118	1.7606125	1.7277082	1.7606125	1.7606125	5
excellent	1.2155508	1.1060514	0.9965587	1.1060514	1.1060514	5
my_first	1.1209985	0.5486933	0.2081031	0.5486933	0.5486933	5
easy	1.0169599	0.9260635	0.8703357	0.9260635	0.9260635	5
perfect	0.9039296	0.6524267	0.510088	0.6524267	0.6524267	5
impress	0.8595221	0.6132996	0.4602385	0.6132996	0.6132996	5
back_for	0.8271636	0.5089204	0.3158973	0.5089204	0.5089204	5
quick	0.8150146	0.6692242	0.5879905	0.6692242	0.6692242	5
fantastic	0.7642072	0.5911542	0.500492	0.5911542	0.5911542	5
within_minute	0.7637819	0.3467163	0.0515449	0.3467163	0.3467163	5
amazing	0.7536879	0.5683715	0.4642053	0.5683715	0.5683715	5
continue	0.7123623	0.4945143	0.3565506	0.4945143	0.4945143	5
and_help	0.6940792	0.3949878	0.2057348	0.3949878	0.3949878	5
thank	0.5552965	0.506089	0.481362	0.506089	0.506089	5
full_refund	0.5384723	0.3863128	0.2904346	0.3863128	0.3863128	5
great_service	0.4513193	0.2482341	0.1315517	0.2482341	0.2482341	5
refundable_hotel	0.448031	0.2503497	0.1319644	0.2503497	0.2503497	5
bit	0.435254	0.1913623	0.0425053	0.1913623	0.1913623	5
had	0.4052053	0.2911124	0.2136364	0.2911124	0.2911124	5
was_able	0.341869	0.2191523	0.1510289	0.2191523	0.2191523	5
none	0.3339257	0.1824334	0.0834002	0.1824334	0.1824334	5
were_verified	0.7820855	0.2170775	0	0.2170775	0.2170775	4
bad_review	0.4358002	0.0438122	0	0.0438122	0.0438122	4
cheap	0.3323283	0.1161286	0	0.1161286	0.1161286	4
had_book	0.1646458	0.0196089	0	0.0196089	0.0196089	4
good	0.1159964	0.0376419	0	0.0376419	0.0376419	4
so_far	0.3159153	0	0	0	0	1
did_so	0.2995969	0	0	0	0	1
and	0.2195096	0	0	0	0	1
comfort	0.1994573	0	0	0	0	1
last_minute	0.162742	0	0	0	0	1
excellent_service	0.155303	0	0	0	0	1
custom	0.1426258	0	0	0	0	1
my_behalf	0.0746737	0	0	0	0	1
overwhelming	0.0629221	0	0	0	0	1
only	0.0615804	0	0	0	0	1
best	0.0542489	0	0	0	0	1
other_companies	0.0490777	0	0	0	0	1
little	0.008604	0	0	0	0	1

Identifying factors of customer satisfaction

After the significant words of customer satisfaction were identified by LASSO, the essential words were ranked by their occurrences using the five-fold cross-validation experiment. The occurrence refers to how many times the words appear in the five experiments. As listed in Table 2, this study only obtained 5, 4, and 1 times of word occurrence frequency following LASSO regulations. If the words with coefficient were not zero showed up more within 5 experiments, it inferred the words were more significant. To diminish subjectivity in word labeling, those words that had similar meanings, purposes, and frequencies were gathered together. This method is simple and objective. Results showed that refunds, promptness, easiness, and assurance were the first-ranked factors placed in the code F1. Bad reviews and cheap were the second-ranked factors placed in the code F2. Excellent service and comparison were the third-ranked factors placed in the code F3. However, experiences were not categorized into a factor because customers showed their experiences with non-meaningful words. Due to the sentiment words only showing gladness and disappointment without meaningful information, it was also not categorized as a factor. Table 3 lists the factors after the words are labeled and ranked based on their occurrences.

DISCUSSION

Refunds, promptness, easiness, and assurance were found as first-ranked factors to OTA customer satisfaction in this study. The refund became a thorny problem to OTAs during the COVID-19 pandemic (*Connor, 2020*; *Piccinelli, Moro & Rita, 2021*). Many airline and hotel customers had to cancel tickets and bookings but some went through complicated refund processes (*Uğur & Akbıyık, 2020*; *Piccinelli, Moro & Rita, 2021*). Customers need an easy and agile process for the booking and refunding process (*Tsang, Lai & Law, 2010*). Promptness is important during the COVID-19 pandemic because travelers can become dissatisfied if the requests are not served within the allowed time (*Lee & Ko, 2021*). Easy process is required by travelers when they requested services, especially during the COVID-19 pandemic (*Foroudi, Tabaghdehi & Marvi, 2021*). Assurance was also found as an important factor for travelers, and it was always during the pandemic as *Uğur & Akbıyık (2020)* stated during the pandemic, travelers want tourism providers to give them assurance services.

Bad reviews and cheap were found as the second-ranked factors in this study. Previous studies suggested that customers' comments either negative or positive are influenced by customer satisfaction (*Berezina et al., 2016*; *Xu, 2020*). This study found negative reviews as the second-ranked factor to customer satisfaction. It is an alert to OTAs that customers' negative comments have greater impacts on potential travelers than those positive messages (*Rianthong, Dumrongsiri & Kohda, 2016*; *Sánchez-Franco, Navarro-García & Rondán-Cataluña, 2019*). Negative comments for hospitality and tourism industries possibly impair OTAs' reputations and block orderings from the existing and future customers during the COVID-19 outbreak (*Luo & Xu, 2021*). Cheap was an important factor for customer satisfaction because most travelers were used to searching for bargain products or services among OTAs during the COVID-19 outbreak (*Nilashi et al., 2022*).

Table 3 Associated factors with customer satisfaction.

Frequency	Code	Factors	Words
5	F1	Refunds	full_refund, refundable_hotel
		Promptness	quick, within_minute
		Easiness	easy
		Assurance	and_help
		Experiences	my_first, never_had, was_able
		Sentiment	thank_you, great, excellent, perfect, impress, fantastic, amazing, great_service, thank, bit, had, none, back_for, continue
4	F2	Bad reviews	bad_review
		Cheap	cheap
		Experiences	were_verified, had_book
		Sentiment	good
1	F3	Excellent service	excellent_service
		Comparison	other_companies
		Experiences	last_minute, custom, only, little, so_far, did_so, and
		Sentiment	comfort, overwhelming, my_behalf, best

Excellent service and comparison were the third-ranked factors. Quality service is always the first priority for customers. During the pandemic, travelers are used to comparing offerings among OTAs and choosing the best one (*Nilashi et al., 2022*). During the pandemic, choosing excellent services with comparing offerings among OTAs became a priority for travelers (*Nilashi et al., 2022*).

Overall, this study contends that external factors other than core services, such as negative reviews and comparison, have an impact on customer satisfaction. These findings differ from those of previous studies (Table 1) which found that only internal factors have a positive influence on customer satisfaction. On the other hand, this study confirms that internal factors have a significant impact on customer satisfaction.

The coronavirus pandemic has influenced industries worldwide and tested companies' capabilities to manage the crisis. It has changed individuals' traveling behavior, OTAs' marketing programs must align with this trend. This study reveals a new set of critical factors to OTA customer satisfaction during the COVID-19 pandemic which informs traveling industries to transform their customer satisfaction's indicators.

CONCLUSION

This study empirically examines the critical factors of customer satisfaction toward online travel agencies when COVID-19 happened in the world. Based on the online customer reviews during the COVID-19 pandemic, a text mining method including the LASSO approach was used to extract the significant factors of customer satisfaction toward OTAs. This approach is feasible to explore extensive issues for travel industries.

During the COVID-19 outbreak, many OTAs have endured great losses from the shortage of orders and faced a bleeding bottom-line of the financial situation. This study helps

OTAs to re-examine their service priorities in order to do trade-off offerings. Regarding the questions of what are the most and critical attributes of customer satisfaction and also the ranking of those attributes. Refunds, promptness, easiness, and assurance were on the first-ranked, followed by bad reviews & cheap in the second-ranked and excellent service & comparison in the third-ranked list. Refunds, bad reviews, assurance, and comparison are ranked as novel factors of customer satisfaction. Understanding the new set of customer satisfaction factors provides insights for OTAs. Managers may place the first-ranked factors to be the top list of their services. Therefore, the generalization of results to other OTAs should be cautious.

Facing the global recession in the tourism industry caused by COVID-19, it is suggested that OTAs redesign competitive offerings to stimulate customer satisfaction during and post-pandemic crises. Second, OTAs should coordinate with tourism suppliers to make easy and fast refund policies with assurance service and procedures for customers. Also, OTAs can re-examine their competitive positions through OCRs, especially good and bad reviews.

Online customer reviews are a valuable source for hospitality and tourism research, their applications are still under-investigated. A limitation of this study is solely collecting OCRs to an OTA from a single review website. To improve the external validity of results, future studies can collect OCRs of multiple online traveling agencies.

ADDITIONAL INFORMATION AND DECLARATIONS

Funding

The authors received no funding for this work.

Competing Interests

The authors declare there are no competing interests.

Author Contributions

- Weishen Wu conceived and designed the experiments, performed the experiments, analyzed the data, authored or reviewed drafts of the paper, and approved the final draft.
- Dalianus Riantama conceived and designed the experiments, performed the experiments, analyzed the data, performed the computation work, prepared figures and/or tables, and approved the final draft.

Data Availability

The following information was supplied regarding data availability:

The data was retrieved from https://www.trustpilot.com/review/www.booking.com and is available as a Supplemental File. The MATLAB code is available as a Supplemental File.

Supplemental Information

Supplemental information for this article can be found online at http://dx.doi.org/10.7717/peerj-cs.850#supplemental-information.

REFERENCES

Atalık Ö, Bakır M, Akan Ş. 2019. The role of in-flight service quality on value for money in business class: a logit model on the airline industry. *Administrative Sciences* **9**:26 DOI 10.3390/admsci9010026.

Berezina K, Bilgihan A, Cobanoglu C, Okumus F. 2016. Understanding satisfied and dissatisfied hotel customers: text mining of online hotel reviews. *Journal of Hospitality Marketing and Management* **25**:1–24 DOI 10.1080/19368623.2015.983631.

Brun I, Rajaobelina L, Ricard L, Amiot T. 2020. Examining the influence of the social dimension of customer experience on trust towards travel agencies: the role of experiential predisposition in a multichannel context. *Tourism Management Perspectives* **34**:100668 DOI 10.1016/j.tmp.2020.100668.

Cai J, Luo J, Wang S, Yang S. 2018. Feature selection in machine learning: a new perspective. *Neurocomputing* **300**:70–79 DOI 10.1016/j.neucom.2017.11.077.

Chang JR, Chen MY, Chen LS, Chien WT. 2020. Recognizing important factors of influencing trust in O2O models: an example of OpenTable. *Soft Computing* **24**:7907–7923 DOI 10.1007/s00500-019-04019-x.

Chang JR, Chen MY, Chen LS, Tseng SC. 2019. Why customers don't revisit in tourism and hospitality industry? *IEEE Access* **7**:146588–146606 DOI 10.1109/ACCESS.2019.2946168.

Chang YW, Hsu PY, Lan YC. 2019. Cooperation and competition between online travel agencies and hotels. *Tourism Management* **71**:187–196 DOI 10.1016/j.tourman.2018.08.026.

Chen K, Zhang Z, Long J, Zhang H. 2016. Turning from TF-IDF to TF-IGM for term weighting in text classification. *Expert Systems with Applications* **66**:245–260 DOI 10.1016/j.eswa.2016.09.009.

Chen MY, Chang JR, Chen LS, Shen EL. 2021. The key successful factors of video and mobile game crowdfunding projects using a lexicon-based feature selection approach. *Journal of Ambient Intelligence and Humanized Computing* 1–19 DOI 10.1007/s12652-021-03146-4.

Chen MC, Hsiao YH, Chang KC, Lin MK. 2019. Applying big data analytics to support Kansei engineering for hotel service development. *Data Technologies and Applications* **53**:33–57 DOI 10.1108/DTA-05-2018-0048.

Cheng M, Jin X. 2019. What do Airbnb users care about? An analysis of online review comments. *International Journal of Hospitality Management* **76**:58–70 DOI 10.1016/j.ijhm.2018.04.004.

Chen CF, Kao YL. 2010. Relationships between process quality, outcome quality, satisfaction, and behavioural intentions for online travel agencies - evidence from Taiwan. *Service Industries Journal* **30**:2081–2092 DOI 10.1080/02642060903191108.

Chiu MC, Lin KZ. 2018. Utilizing text mining and Kansei Engineering to support data-driven design automation at conceptual design stage. *Advanced Engineering Informatics* **38**:826–839 DOI 10.1016/j.aei.2018.11.002.

Connor P. 2020. More than nine-in-ten people worldwide live in countries with travel restrictions amid COVID-19. *Available at https://www.pewresearch.org/fact-tank/2020/04/01/more-than-nine-in-ten-people-worldwide-live-in-countries-with-travel-restrictions-amid-covid-19/*.

Cui F, Lin D, Qu H. 2018. The impact of perceived security and consumer innovativeness on e-loyalty in online travel shopping. *Journal of Travel and Tourism Marketing* **35**:819–834 DOI 10.1080/10548408.2017.1422452.

Dastjerdi AR, Foroghi D, Kiani GH. 2019. Detecting manager's fraud risk using text analysis: evidence from Iran. *Journal of Applied Accounting Research* **20**:154–171 DOI 10.1108/JAAR-01-2018-0016.

Davras Ö, Caber M. 2019. Analysis of hotel services by their symmetric and asymmetric effects on overall customer satisfaction: a comparison of market segments. *International Journal of Hospitality Management* **81**:83–93 DOI 10.1016/j.ijhm.2019.03.003.

Evans JR, Mathur A. 2018. The value of online surveys: a look back and a look ahead. *Internet Research* **28**:854–887 DOI 10.1108/IntR-03-2018-0089.

Farhadloo M, Patterson RA, Rolland E. 2016. Modeling customer satisfaction from unstructured data using a Bayesian approach. *Decision Support Systems* **90**:1–11 DOI 10.1016/j.dss.2016.06.010.

Foroudi P, Tabaghdehi SAH, Marvi R. 2021. The gloom of the COVID-19 shock in the hospitality industry: a study of consumer risk perception and adaptive belief in the dark cloud of a pandemic. *International Journal of Hospitality Management* **92**:102717 DOI 10.1016/j.ijhm.2020.102717.

Gandomi A, Haider M. 2015. Beyond the hype: big data concepts, methods, and analytics. *International Journal of Information Management* **35**:137–144 DOI 10.1016/j.ijinfomgt.2014.10.007.

Gopi B, Samat N. 2020. The influence of food trucks' service quality on customer satisfaction and its impact toward customer loyalty. *British Food Journal* **122**:3213–3226 DOI 10.1108/BFJ-02-2020-0110.

Guo L, Vargo CJ, Pan Z, Ding W, Ishwar P. 2016. Big social data analytics in journalism and mass communication. *Journalism & Mass Communication Quarterly* **93**:332–359 DOI 10.1177/1077699016639231.

Guyon I, Elisseeff A. 2003. An introduction to variable and feature selection. *Journal of Machine Learning Research* **3**:1157–1182 DOI 10.1162/153244303322753616.

Guyon I, Elisseefl A. 2006. Feature extraction: an introduction to feature extraction. In: Kacprzyk J, ed. *Studies in fuzziness and soft computing*. New York: Springer Berlin Heidelberg, 1–25 DOI 10.1007/978-3-540-35488-8_1.

Ha HY, Janda S. 2016. The evolution of expectations of and attitudes toward online travel agencies over time. *Journal of Travel and Tourism Marketing* **33**:966–980 DOI 10.1080/10548408.2015.1075457.

Hao F, Xiao Q, Chon K. 2020. COVID-19 and China's hotel industry: impacts, a disaster management framework, and post-pandemic agenda. *International Journal of Hospitality Management* **90**:102636 DOI 10.1016/j.ijhm.2020.102636.

Hao JX, Yu Y, Law R, Fong DKC. 2015. A genetic algorithm-based learning approach to understand customer satisfaction with OTA websites. *Tourism Management* **48**:231–241 DOI 10.1016/j.tourman.2014.11.009.

Hernández-Ortega B. 2018. Don't believe strangers: online consumer reviews and the role of social psychological distance. *Information and Management* **55**:31–50 DOI 10.1016/j.im.2017.03.007.

Hlee S, Lee H, Koo C, Chung N. 2020. Will the relevance of review language and destination attractions be helpful? A data-driven approach. *Journal of Vacation Marketing* **27**:61–81 DOI 10.1177/1356766720950356.

Hong JW, Park SB. 2019. The identification of marketing performance using text mining of airline review data. *Mobile Information Systems* **2019**:1–8 DOI 10.1155/2019/1790429.

Hou Z, Cui F, Meng Y, Lian T, Yu C. 2019. Opinion mining from online travel reviews: a comparative analysis of Chinese major OTAs using semantic association analysis. *Tourism Management* **74**:276–289 DOI 10.1016/j.tourman.2019.03.009.

Hsu CL, Chang KC, Chen MC. 2012. The impact of website quality on customer satisfaction and purchase intention: Perceived playfulness and perceived flow as mediators. *Information Systems and e-Business Management* **10**:549–570 DOI 10.1007/s10257-011-0181-5.

Huang J, Wang H, Kochenberger G. 2017. Distressed Chinese firm prediction with discretized data. *Management Decision* **55**:786–807 DOI 10.1108/MD-08-2016-0546.

Jia SS. 2018. Behind the ratings: text mining of restaurant customers' online reviews. *International Journal of Market Research* **60**:561–572 DOI 10.1177/1470785317752048.

Joung J, Kim KH, Kim K. 2021. Data-driven approach to dual service failure monitoring from negative online reviews: managerial perspective. *SAGE Open* **11**:1–14 DOI 10.1177/2158244020988249.

Ju Y, Back KJ, Choi Y, Lee JS. 2019. Exploring Airbnb service quality attributes and their asymmetric effects on customer satisfaction. *International Journal of Hospitality Management* **77**:342–352 DOI 10.1016/j.ijhm.2018.07.014.

Kim J, Franklin D, Phillips M, Hwang E. 2020. Online travel agency price presentation: examining the influence of price dispersion on travelers' hotel preference. *Journal of Travel Research* **59**:704–721 DOI 10.1177/0047287519857159.

Kim WG, Lee HY. 2004. Comparison of web service quality between online travel agencies and online travel suppliers. *Journal of Travel & Tourism Marketing* **17**:105–116 DOI 10.1300/J073v17n02_09.

Kohavi R, John GH. 1998. Feature extraction, construction and selection: the wrapper approach. In: Liu H, Motoda H, eds. *The springer international series in engineering and computer science*. New York: Springer Science+Business Media, 33–50 DOI 10.1007/978-1-4615-5725-8_3.

Kwok JT, Zhou ZH, Xu L. 2015. Springer handbook of computational intelligence: machine learning. In: Kacprzyk J, Pedrycz W, eds. *Springer handbooks*. New York: Springer Berlin Heidelberg, 495–522 DOI 10.1007/978-3-662-43505-2.

Lal TN, Chapelle O, Western J, Elisseeff A. 2006. Feature extraction: embedded methods. In: Guyon I, Nikravesh M, Gunn S, Zadeh LA, eds. *Studies in fuzziness and soft computing*. New York: Springer Berlin Heidelberg, 137–165 DOI 10.1007/978-3-540-35488-8_6.

Lee W, Ko YD. 2021. Operation policy of multi-capacity logistic robots in hotel industry. *International Journal of Contemporary Hospitality Management* **33**:1482–1506 DOI 10.1108/IJCHM-05-2020-0372.

Lien CH, Wen MJ, Wu CC. 2011. Investigating the relationships among E-service quality, perceived value, satisfaction, and behavioral intentions in Taiwanese online shopping. *Asia Pacific Management Review* **16**:211–223 DOI 10.6126/APMR.2011.16.3.01.

Lim H, Kim DW. 2020. MFC: initialization method for multi-label feature selection based on conditional mutual information. *Neurocomputing* **382**:40–51 DOI 10.1016/j.neucom.2019.11.071.

Lockyer T. 2005. The perceived importance of price as one hotel selection dimension. *Tourism Management* **26**:529–537 DOI 10.1016/j.tourman.2004.03.009.

Long Y, Shi P. 2017. Pricing strategies of tour operator and online travel agency based on cooperation to achieve O2O model. *Tourism Management* **62**:302–311 DOI 10.1016/j.tourman.2017.05.002.

Luo Y, Xu X. 2021. Comparative study of deep learning models for analyzing online restaurant reviews in the era of the COVID-19 pandemic. *International Journal of Hospitality Management* **94**:102849 DOI 10.1016/j.ijhm.2020.102849.

Makarov I, Gerasimova O, Sulimov P, Zhukov LE. 2019. Dual network embedding for representing research interests in the link prediction problem on co-authorship networks. *PeerJ Computer Science* **5(9)**:e172 DOI 10.7717/peerj-cs.172.

Mudambi SM, Schuff D. 2010. What makes a helpful online review? A study of customer reviews on amazon.com. *MIS Quarterly: Management Information Systems* **34**:185–200 DOI 10.2307/20721420.

Nilashi M, Abumalloh RA, Minaei-Bidgoli B, Abdu Zogaan W, Alhargan A, Mohd S, Syed Azhar SNF, Asadi S, Samad S. 2022. Revealing travellers' satisfaction during COVID-19 outbreak: moderating role of service quality. *Journal of Retailing and Consumer Services* **64**:102783 DOI 10.1016/j.jretconser.2021.102783.

Nunkoo R, Teeroovengadum V, Ringle CM, Sunnassee V. 2020. Service quality and customer satisfaction: the moderating effects of hotel star rating. *International Journal of Hospitality Management* **91**:102414 DOI 10.1016/j.ijhm.2019.102414.

Piccinelli S, Moro S, Rita P. 2021. Air-travelers' concerns emerging from online comments during the COVID-19 outbreak. *Tourism Management* **85**:104313 DOI 10.1016/j.tourman.2021.104313.

Pereira HG, de Fátima Salgueiro M, Rita P. 2017. Online determinants of e-customer satisfaction: application to website purchases in tourism. *Service Business* **11**:375–403 DOI 10.1007/s11628-016-0313-6.

Prassida GF, Hsu P-Y, Chang Y-W. 2021. Understanding how O2O service synergies drive customer continuance intention: a study of OTAs and hotels. *Asia Pacific Journal of Tourism Research* **26**:1139–1155 DOI 10.1080/10941665.2021.1952461.

Rajaobelina L. 2018. The impact of customer experience on relationship quality with travel agencies in a multichannel environment. *Journal of Travel Research* **57**:206–217 DOI 10.1177/0047287516688565.

Ramos J. 2003. Using TF-IDF to determine word relevance in document queries. In: *Proceedings of the first instructional conference on machine learning. Washington D.C., USA, 21-24 August*. 12–14.

Rianthong N, Dumrongsiri A, Kohda Y. 2016. Improving the multidimensional sequencing of hotel rooms on an online travel agency web site. *Electronic Commerce Research and Applications* **17**:74–86 DOI 10.1016/j.elerap.2016.03.002.

Sánchez-Franco MJ, Navarro-García A, Rondán-Cataluña FJ. 2019. A naive Bayes strategy for classifying customer satisfaction: a study based on online reviews of hospitality services. *Journal of Business Research* **101**:499–506 DOI 10.1016/j.jbusres.2018.12.051.

Sant'Anna LR, Caldeira JF, Filomena TP. 2020. Lasso-based index tracking and statistical arbitrage long-short strategies. *North American Journal of Economics and Finance* **51**:101055 DOI 10.1016/j.najef.2019.101055.

Sebastiani F. 2002. Machine learning in automated text categorization. *ACM Computing Surveys* **34**:1–47 DOI 10.1145/505282.505283.

Sezgen E, Mason KJ, Mayer R. 2019. Voice of airline passenger: a text mining approach to understand customer satisfaction. *Journal of Air Transport Management* **77**:65–74 DOI 10.1016/j.jairtraman.2019.04.001.

Sharma A, Sharma S, Chaudhary M. 2020. Are small travel agencies ready for digital marketing? Views of travel agency managers. *Tourism Management* **79**:104078 DOI 10.1016/j.tourman.2020.104078.

Sigala M. 2020. Tourism and COVID-19: impacts and implications for advancing and resetting industry and research. *Journal of Business Research* **117**:312–321 DOI 10.1016/j.jbusres.2020.06.015.

Singh R, Söderlund M. 2020. Extending the experience construct: an examination of online grocery shopping. *European Journal of Marketing* **54**:2419–2446 DOI 10.1108/EJM-06-2019-0536.

Situmeang F, de Boer N, Zhang A. 2020. Looking beyond the stars: a description of text mining technique to extract latent dimensions from online product reviews. *International Journal of Market Research* **62**:195–215 DOI 10.1177/1470785319863619.

Sthapit E, Del Chiappa G, Coudounaris DN, Bjork P. 2020. Determinants of the continuance intention of Airbnb users: consumption values, co-creation, information overload and satisfaction. *Tourism Review* **75**:511–531 DOI 10.1108/TR-03-2019-0111.

Sullivan D. 2001. *Document warehousing and text mining: techniques for improving business operations, marketing, and sales*. New York: John Wiley & Sons, Inc.

Swiniarski RW, Skowron A. 2003. Rough set methods in feature selection and recognition. *Pattern Recognition Letters* **24**:833–849 DOI 10.1016/S0167-8655(02)00196-4.

Tahanisaz S, Shokuhyar S. 2020. Evaluation of passenger satisfaction with service quality: A consecutive method applied to the airline industry. *Journal of Air Transport Management* **83**:101764 DOI 10.1016/j.jairtraman.2020.101764.

Tao S, Kim HS. 2019. Cruising in Asia: what can we dig from online cruiser reviews to understand their experience and satisfaction. *Asia Pacific Journal of Tourism Research* **24**:514–528 DOI 10.1080/10941665.2019.1591473.

Tibshirani R. 1996. Regression shrinkage and selection via the Lasso. *Journal of the Royal Statistical Society: Series B (Methodological)* **58**:267–288 DOI 10.1111/j.2517-6161.1996.tb02080.x.

Ting C-W, Chen M-S, Lee C-L. 2013. E-satisfaction and post-purchase behaviour of online travel product shopping. *Journal of Statistics and Management Systems* **16**:223–240 DOI 10.1080/09720510.2013.821337.

Trefis Team, Great Speculations. 2019. Booking holdings and Expedia are both growing steadily, but in very different ways. Forbes. *Available at https://www.forbes.com/sites/greatspeculations/2019/08/26/booking-holdings-and-expedia-are-both-growing-steadily-but-in-very-different-ways/#36eb4be71f04* (accessed on 12 September 2020).

Tsang NKF, Lai MTH, Law R. 2010. Measuring E-service quality for online travel agencies. *Journal of Travel and Tourism Marketing* **27**:306–323 DOI 10.1080/10548401003744743.

Uğur NG, Akbıyık A. 2020. Impacts of COVID-19 on global tourism industry: a cross-regional comparison. *Tourism Management Perspectives* **36**:100744 DOI 10.1016/j.tmp.2020.100744.

Wang T. 2021. A combined model for short-term wind speed forecasting based on empirical mode decomposition, feature selection, support vector regression and cross-validated lasso. *PeerJ Computer Science* **7**:e732 DOI 10.7717/peerj-cs.732.

Wang Y, Zhang M, Tse YK, Chan HK. 2020. Unpacking the impact of social media analytics on customer satisfaction: do external stakeholder characteristics matter? *International Journal of Operations and Production Management* **40**:647–669 DOI 10.1108/IJOPM-04-2019-0331.

Xu X. 2020. Examining an asymmetric effect between online customer reviews emphasis and overall satisfaction determinants. *Journal of Business Research* **106**:196–210 DOI 10.1016/j.jbusres.2018.07.022.

Xu X, Li Y. 2016. Examining key drivers of traveler dissatisfaction with airline service failures: a text mining approach. *Journal of Supply Chain and Operations Management* **14**:30–50.

Yang D, Kleissl J, Gueymard CA, Pedro HTC, Coimbra CFM. 2018. History and trends in solar irradiance and PV power forecasting: a preliminary assessment and review using text mining. *Solar Energy* **168**:60–101 DOI 10.1016/j.solener.2017.11.023.

Yoon Y, Uysal M. 2005. An examination of the effects of motivation and satisfaction on destination loyalty: a structural model. *Tourism Management* **26**:45–56 DOI 10.1016/j.tourman.2003.08.016.

Yüksel A, Yüksel F. 2001. The expectancy-disconfirmation paradigm: a critique. *Journal of Hospitality & Tourism Research* **25**:107–131 DOI 10.1177/109634800102500201.

Zhang CH, Huang J. 2008. The sparsity and bias of the lasso selection in high-dimensional linear regression. *Annals of Statistics* **36**:1567–1594 DOI 10.1214/07-AOS520.

Zhang K, Hou Y, Li G. 2020. Threat of infectious disease during an outbreak: Influence on tourists' emotional responses to disadvantaged price inequality. *Annals of Tourism Research* **84**:102993 DOI 10.1016/j.annals.2020.102993.

Zhao P, Yu B. 2006. On model selection consistency of Lasso. *Journal of Machine Learning Research* **7**:2541–2563.

Zhou Y, Xue Y. 2020. ACRank: a multi-evidence text-mining model for alliance discovery from news articles. *Information Technology and People* **33**:1357–1380 DOI 10.1108/ITP-06-2018-0272.

Zinko R, Furner CP, De Burgh-Woodman H, Johnson P, Sluhan A. 2021. The addition of images to eWOM in the travel industry: an examination of hotels, cruise ships and fast food reviews. *Journal of Theoretical and Applied Electronic Commerce Research* **16**:525–541 DOI 10.3390/jtaer16030032.