Decision Support Systems 107 (2018) 52–63

Contents lists available at ScienceDirect

Decision Support Systems

journal homepage: www.elsevier.com/locate/dss

Disentangling consumer recommendations: Explaining and predicting airline recommendations based on online reviews

Michael Siering [a], Amit V. Deokar [b,*], Christian Janze [a]

[a] Goethe University Frankfurt, Theodor-W.-Adorno-Platz 4, 60323 Frankfurt, Germany
[b] University of Massachusetts Lowell, Robert J. Manning School of Business, One University Avenue, Lowell, MA 01854, United States

ARTICLE INFO

Article history:
Received 12 June 2017
Received in revised form 7 November 2017
Accepted 6 January 2018
Available online 11 January 2018

Keywords:
Consumer recommendation
Promoter score
Online review
Sentiment analysis
Data mining

ABSTRACT

Consumer recommendations of products and services are important performance indicators for organizations to gain feedback on their offerings. Furthermore, they are important for prospective customers to learn from prior consumer experiences. In this study, we focus on user-generated content, in particular online reviews, to investigate which service aspects are evaluated by consumers and how these factors explain a consumer's recommendation. Further, we investigate how recommendations can be predicted automatically based on such user-driven responses. We disentangle the recommendation decision by performing explanatory and predictive analyses focusing on a sample of airline reviews. We identify core and augmented service aspects expressed in the online review. We then show that service aspect-specific sentiment indicators drive the decision to recommend an airline and that these factors can be incorporated in a predictive model using data mining techniques. We also find that the business model of an airline being reviewed, whether low cost or full service, is also an applicable consideration. Our results are highly relevant for practitioners to analyze and act on consumer feedback in a prompt manner, along with the ability of gaining a deeper understanding of the service from multiple aspects. Also, potential travelers can benefit from this approach by getting an aggregated view on service quality.

1. Introduction

Sensing and understanding consumer perceptions is important for corporations to deal with positive and negative consumer feedback as well as to engage prospective consumers with their product or service offerings. Previous research indicates that performance metrics in form of promoter scores are invaluable to corporations to estimate consumer satisfaction [1]. For determining such metrics, corporations take into account whether consumers recommend a specific service or not in order to learn whether consumers are satisfied.

Nowadays, consumers increasingly rely on electronic word-of-mouth as a mechanism to share their experiences of their own volition and express their satisfaction in experiencing a product or service [2, 3]. Online reviews usually appear as freeform text and often embed a multitude of experiences and opinions regarding different aspects of the reviewed service. However, not all reviewers directly express by means of a standardized field whether they recommend the evaluated service which makes it hard for corporations to calculate promoter scores.

The availability of rich information buried in the reviews' texts provides an opportunity to enhance the decision-making capabilities of stakeholders in this context, i.e., prospective consumers as well as service providers. First, examining whether the textual contents of each review can uncover reviewers' assessment of recommending the service or not, which can be informative itself. Second, the ability to disentangle the recommendation decision to examine how specific service aspects are being evaluated by prior consumers and how they drive the overall recommendation decision can further inform the stakeholders. A decision support approach analyzing distinct service aspects and inferring recommendations expressed in online reviews is thus needed to address these relevant issues.

The vital role of online reviews as a key data source in tourism has also been shown in recent studies which focus on the use of online reviews and their role in influencing consumer behavior and decision-making [4, 5]. Furthermore, especially the helpfulness of online reviews has been analyzed [6–9]. Nevertheless, despite these valuable and interesting aspects of online reviews, the recommendation decision and its influencing factors, as key aspects of online reviews, have not been addressed so far.

This study focuses on analyzing and inferring recommendations expressed by reviewers in online reviews in tourism services, particularly airline services, from multiple perspectives. First, we examine which service aspects are expressed in the textual contents of airline

[*] Corresponding author.
E-mail addresses: siering@wiwi.uni-frankfurt.de (M. Siering), amit_deokar@uml.edu (A.V. Deokar), janze@wiwi.uni-frankfurt.de (C. Janze).

https://doi.org/10.1016/j.dss.2018.01.002
0167-9236/© 2018 Elsevier B.V. All rights reserved.

reviews, where we specifically delve into analyzing the role of core and augmented service aspects [10, 11]. We also investigate whether reviewers' recommendations of a service can be explained through service aspects extracted from online reviews. Drawing on theoretical foundations of accessibility-diagnosticity model and multiple pathway anchoring and adjustment model, we posit that services such as airline services are perceived and assessed based on different aspects of the service, and that only a few prominent service aspects play a key role in collectively forming an overall assessment about the service based on which the recommendation decision is reached. This explanatory analysis is particularly useful to service providers in understanding which service aspects are influencing consumer word-of-mouth promotion through online reviews. Second, the study builds on the results from the previous analysis to examine whether reviewers' recommendations of a service can be predicted using contents expressed in online reviews. In building the predictive models, overall content-based sentiments as well as service aspect-specific sentiments are investigated and compared to a classic text mining approach based on a bag-of-words model. The analysis is relevant for both prospective consumers as well as service providers from a decision support standpoint. Consumers can make informed decisions using the predicted recommendation scores, while service providers can obtain performance indicators to be integrated in managerial dashboards for service management. Third, the study also analyzes the role of service business models in the explanation and prediction of service recommendations from online reviews. Given that different service providers have distinct business models such as low-cost carriers versus full-service network carriers in the airline industry, it may be expected that travelers have different expectations about the service, which will likely inform their review and eventual recommendation. The analysis sheds light on these differences in service models.

The remainder of this paper is structured as follows. Section 2 provides an overview on the background of this study, taking into account the recommendation decision as well as the role of online reviews. In this context, the theoretical models underlying our rationale are presented as well. Next, Section 3 focuses on the research methodology applied, including the explanatory and predictive analyses performed. Section 4 presents the results of our analyses. Section 5 discusses issues related to our findings, and Section 6 concludes the article.

2. Background and research questions

2.1. Consumer recommendations

For corporations, customer feedback is essential, not only for driving business performance and growth, but also for product and service innovation as well as for improving customer experience. Managers traditionally rely on customer feedback metrics including average customer satisfaction, top-2-box customer satisfaction, proportion of customers complaining, repurchase likelihood of customers as well as word-of-mouth communication such as number of recommendations or promoters [12].

In the field of consumer recommendations, Reichheld [1] proposed the notion of word-of-mouth product or service recommendations, aggregated as the *Net Promoter Score* (NPS), to be the single most reliable metric to predict a business' ability to grow, particularly in relation to other aforementioned customer feedback metrics. NPS is computed simply as the difference in percentage between promoters and detractors. This proposition has received much scrutiny by academics and practitioners alike over the past decade. Previous research has found that a growth in NPS is correlated with a growth in business [13]. In support of Reichheld's arguments, it was also found that promoter scores are one of the best customer feedback metrics in predicting customer retention [14]. In industry, promoter scores have been leveraged to gain insight into the customer base and in turn to drive market share growth [15]. Thus, programs based on promoter scores are particularly effective in underscoring the overall strategy of listening to customers, and substantiating changes transparently by attributing them to relevant testimonies [15].

Given the importance of customer engagement, loyalty, and feedback, as reflected in promoter scores, it is also imperative to understand the factors that influence positive word-of-mouth, i.e. service or product recommendation. As consumers nowadays express their experiences within online reviews, it is essential for corporations to monitor this form of electronic word-of-mouth in order to understand consumers and to identify the aspects influencing the recommendation decision. Nevertheless, many businesses struggle to utilize online reviews to create business value [16]. This is because most reviews are not *directly* tied to a service or product recommendation score, i.e., the issue of whether a customer recommends the service to another potential customer. Thus, a research gap exists in the area of automatically extracting information from online reviews to accurately derive *indirect* recommendation intentions, disentangling the recommendation decision into different service aspects expressed, and to ultimately incorporate them within a promoter score.

Notably, promoter scores based on consumer recommendations differ from typical online recommendations considering the major variable of interest: Whereas the rich research stream on online recommendations focuses on the question of how a specific product or service can be selected from *a set of products or services* and be recommended to a consumer, the research stream on consumer recommendations investigates whether consumers recommend a *specific product or service of interest*. Whereas numerous studies in the field of product recommendation agents providing online recommendations do exist [17–19], there is a research gap in the field of extracting consumer recommendations from online reviews.

2.2. Online reviews

Consumers increasingly rely on online reviews as an important information source to base their purchase decisions on [20]. In this context, previous research indicates that online reviews can be seen as information cues during the different phases of the purchase decision making process [21] and that online reviews consequently influence the demand for the services reviewed [22]. Furthermore, online reviews are of high value for online retailers as they attract consumers who then, in a next step, might also potentially purchase the reviewed service [8].

There are different research streams focusing on this type of electronic word of mouth. Extant literature indicates a variety of studies relating online reviews to sales figures, whereas these studies report a significant influence of online reviews on the corresponding demand related to a specific service [23, 24]. Another stream of research focuses on the question of what makes online reviews helpful and credible. In this context, different characteristics like a review's depth or specific review contents [25–27], the presentation of the reviews in terms of their order [28] or community membership [29] have been shown to be relevant. From a practical standpoint, determinants of the factors influencing online review helpfulness can be used to improve customer-centric product strategies [30]. Yet another research stream focuses on the reviewers contributing online reviews and mainly investigates the factors that motivate reviewers to contribute content on social commerce platforms [31, 32].

Studies specifically focusing on tourism show that travelers make use of user-generated content such as online reviews as an information source prior to making purchase decisions. Recent surveys among travelers suggest that a total of 20–45% of travelers use user-generated reviews to inform and/or guide their decision making processes ex-ante and a total of 5–30% to share their experiences ex-post [33, 34]. Furthermore, Xiang, Schwartz, Gerdes, and Uysal [35] have studied the nature of hotel guest experiences expressed in customer reviews and examine the role of salient aspects of hotel guest experience in explaining guest satisfaction. The study, however, does not study the sentiment of the service aspects and their role as predictors of consumer recommendation. In the same domain, Rhee and Yang [4] examine the varied importance of hotel service aspects in different hotel segments. While these studies provide valuable insights regarding the role of online reviews in the field of consumer purchase decision making, an important aspect of the consumer recommendation of services in relation to online reviews has not received much attention: An examination of drivers of a recommendation decision and the automatic extraction of the overall recommendation decision from unstructured user-generated online reviews can be valuable to various stakeholders in the tourism industry.

2.3. Text analytics of user generated content

Text analytics or text mining techniques allow for the analysis of unstructured text documents, such as online reviews, to extract meaningful information pieces and derive structured variables for subsequent analyses. Text analytics includes several focus areas such as search and information retrieval, document classification, document clustering, web mining, information extraction, natural language processing, and concept extraction [36]. This study focuses on a text mining approach, referred to as 'text data mining' by Hotho, Nürnberger, and Paaß [37] that entails using information extraction and text pre-processing steps in order to extract data from text, and subsequently applying data mining algorithms on the extracted data.

Bag-of-words is a standard technique used in text data mining applications (e.g., [38]) that relies on the frequency of occurrence of words in a collection of documents (e.g., online reviews) to derive structured data from the text. In this technique, the specific words in the review text are used as features to represent the review and their weight is most commonly determined by constructing term frequency – inverse document frequency (TF-IDF) matrix [39]. This matrix is then used for applying data mining techniques like classification.

With the widespread adoption of social media, sentiment analysis or opinion mining, has emerged as a novel area within text analytics [40]. Sentiment analysis of user-generated content, such as online reviews, may be conducted at document-level (coarse) or sentence-level (granular) when the entire review or each sentence in the review refers to a specific entity, e.g., airline service experience. Another form of sentiment analysis called *aspect-based sentiment analysis* is appropriate when several different aspects of a service are discussed within a given review and the goal is to tease out the opinions regarding each service aspect [40]. A review of the related literature suggests that studies utilizing aspect-based sentiment analysis techniques are limited to a few instances within the e-commerce domain concerning the extraction and examination of product features [41, 42]. However, to the best of our knowledge, this study is a first such application in the context of tourism services, particularly airline services, with the objective of explaining and predicting consumer recommendations from unstructured user-generated online reviews.

2.4. Research questions

We describe theoretical underpinnings that drive our approach in creating explanatory and predictive models for consumer recommendation scores of products or services. Extant research in online ratings has focused on how consumers perceive online ratings and use them in decision-making in using products or services [43, 44]. In a distinct yet analogous manner, we posit that reviewers themselves are also involved in a decision-making process, one that is aimed at assimilating their own experiences with the service at hand in the form of a review, analyzing the service with the view of whether it is worth recommending to someone else.

The *accessibility-diagnosticity* (AD) model [45, 46] explains how people form attitudes that guide behavior (or proximate determinants of behavior such as judgments in their decision-making process) based on relative accessibility and diagnosticity characteristics of inputs. In the case of online reviews, the service experience generates sentiments that serve as a set of inputs that are based on different aspects of that service. For instance, in evaluating a restaurant, reviewers may rely on their experience with various aspects of the service such as ambience, freshness of food, and so forth. Similarly, in evaluating their experience with airlines, customers may rely on service-specific aspects such as seat comfort, inflight entertainment, and so forth. Drawing on the AD model, reviewers' judgment of a service (in terms of their recommendation) is likely to be influenced by their sentiments formed in relation to individual service-specific aspects that are relatively accessible to them, and therefore, come to their mind readily at the time of recommendation formation. Further, customers' sentiments about specific aspects of the service that are perceived to be relevant or diagnostic are likely to influence their recommendations [46].

The *Multiple Pathway Anchoring and Adjustment* (MPAA) model [47] allows to explain the process of the recommendation formation further in terms of how sentiments regarding specific aspects of a service are formed. The MPAA model underscores the idea of multiple pathways to sentiment or attitude formation, including outside-in (object-centered) and inside-out (person-centered) pathways. In the case of online reviews for a service, on one hand, the experience with a certain aspect of a service provides the object-centered pathway to sentiment generation regarding that service aspect. On the other hand, personal factors that customers experience due to the specific context, situation or personal disposition provide the person-centered pathway to their sentiment generation. Lynch [48] elaborates on the complementary nature of the AD and the MPAA models. The MPAA model delves in the attitude formation process, whereas the AD model is silent on how customers may form sentiment regarding a certain service aspect. In a manner similar to the diagnosticity mechanism in the AD model, the MPAA model suggests representational sufficiency as a mechanism of how customers likely assess sentiments regarding multiple aspects of a service when arriving at an overall judgment such as whether the service should be recommended. Essentially, sentiment regarding service-aspects that are distinctively positive or negative than others are said to be diagnostic in forming the overall service recommendation.

We argue that most services such as hotel stays or flights can be dissected in terms of service aspects that together make up the service experience. For example, a hotel service may be characterized with aspects such as value, location, sleep quality, rooms, cleanliness, and service [4]. The AD and MPAA theoretical models together suggest that online reviews implicitly capture the sentiment expressed by the consumers along service aspects which are accessible and diagnostic. Thus, an online review of a service can be said to express a reviewer's synthesized mental model of one or more service aspects that were perceived as salient during the service experience. Further, based on the MPAA model, each of the service aspects is a possible pathway to express his sentiment or opinion about the service. As such, considering multiple aspect-oriented sentiments and assessments is important to explaining the overall recommendation of the service. The online review, as a holistic unit, may express a certain sentiment, which may be segmented into sentiments regarding specific service aspects.

Thus, we posit that the recommendation about the overall service collectively expresses sentiments of individual service aspects. In our study, we first focus on gaining an understanding about which service aspects are expressed. Towards that end, we differentiate between core and augmented service aspects. In this context, core service aspects represent all aspects related to the basic service and the basic customer benefit received, whereas augmented service aspects encompass all aspects which are facilitating or ancillary to the core service, i.e. aspects where the company can differentiate itself from others and that consequently "surround" the service [10, 11]. Further, we investigate whether they can explain reviewers' choice to recommend the service to other consumers or not. Drawing on the theories discussed, we investigate:

Research Question 1a (RQ1a). Which core and augmented service aspects are expressed in online reviews?

Research Question 1b (RQ1b). Are sentiments about service aspects extracted from online reviews valuable in explaining reviewers' recommendations of a service, and what is the role of core as well as augmented service aspects?

Building on the knowledge derived about the aspects that contribute to service recommendations, we are interested in investigating the predictive nature of these service aspects. From a practical standpoint, potential consumers can be presented with predicted recommendation scores along with the reviews to assist in their decision-making process. Also, it is highly relevant for service providers and intermediaries to potentially leverage textual contents of online reviews to derive corporate performance indicators, ultimately to be incorporated in performance dashboards for managerial decision-making. In regard to predicting customer recommendations, we are also interested in learning about the predictive ability of the sentiment expressed regarding different aspects of the service, also compared to review text in a generic sense based on the content itself (i.e., represented by the classical text mining approach based on the bag-of-words model). Thus, the next research question inquires about the predictive power of the textual contents of a review in predicting a service recommendation, particularly in the absence of explicit recommendation information.

Research Question 2 (RQ2). Is sentiment expressed about service aspects valuable in order to predict reviewers' recommendations of a service, and what is the predictive power compared to a classical text mining approach?

Finally, we are also interested in finding whether the results vary based on the type of business model adopted, such as *low cost* model and *full service* model. Over the past two decades, low cost carriers (LCCs) and full service network carriers (FSNCs) have been noted as distinct business models within the airline service industry [49, 50]. On one hand, LCCs are broadly characterized as airlines having low operating costs and offering "no frills" service experience with a lower baseline ticket cost, but adding substantive charges for features like bags, seat selection, in-flight entertainment, and other amenities. On the other hand, FSNCs are characterized as mega-brand airlines with global networks and offering amenities such as club lounges, first class cabins, and more. Travelers are likely to have different expectations from a service experience with a LCC compared to a FSNC [51]. Accordingly, travelers' may be expected to express sentiments more about different service aspects in these two categories based on both how salient a service aspect was in their service experience and its alignment with initial expectations, which might therefore also influence the recommendation. Thus, we investigate:

Research Question 3 (RQ3). Does the type of the service (e.g., full service versus low-cost service) impact the explanation and prediction of service recommendations?

Using data from an online review platform for airlines, we address these research questions by systematically analyzing the relationship between textual contents of online reviews and consumer recommendations, and demonstrating that the approach can be applied to infer consumer recommendations from online reviews.

3. Research methodology

3.1. Research process

To investigate the three research questions, focused on examining which service aspects are expressed in online reviews and whether a reviewer's recommendation can be explained by sentiment on core and augmented service aspects expressed in the review (i.e., RQ1a/b), whether the review's contents have predictive power to infer the recommendation (i.e., RQ2), and whether the type of business model impacts the explanation and prediction of service recommendations (i.e., RQ3), we adapt the structured knowledge discovery process proposed by Han and Kamber [52].

At the beginning, we select and transform an appropriate dataset and identify and extract the specific service aspects (RQ1a). Following this, we perform an explanatory analysis to understand the significance of the sentiment expressed related to the service aspects discussed (RQ1b). This is also used to determine the variables to be used as predictors in the subsequent predictive analysis that investigates RQ2. In predictive modeling of the data, we evaluate the applicability of different machine learning algorithms to infer the consumer recommendation of the service. We evaluate the different predictive models using a stratified 10-fold cross-validation approach to draw conclusions about the applicability of the proposed aspect-oriented sentiment variables as well as a classic text mining approach. Last, but not least, we conduct the explanatory and predictive analyses for distinct service types, addressing RQ3. Fig. 1 provides an overview of our research process. The specific analyses have been performed with Stata and RapidMiner.

3.2. Data acquisition

In this study, we rely on random sampling of airline reviews that are published on the airline evaluation platform airlinequality.com. This platform represents a comprehensive information source where travelers can evaluate and research the quality of different airlines and their services offered. The platform offers the ability for a traveler to post detailed textual review regarding a flight experience with an airline. Furthermore, users can publish their evaluation of the airline, information regarding the flight purchased (e.g. the date as well as the route taken) as well as evaluations of different aspects of the flight experience (e.g., seat comfort, food service, etc.). In addition, the platform provides an option for the travelers to explicitly respond to the question of whether they would recommend the airline to other travelers or not.

In the following, three samples, each consisting of 1000 airlines reviews and originating from the same platform are considered. The first sample is a random sample from the airline review platform, covering in total 195 different airlines. The other two samples are stratified random samples based on the type of airline, namely (a) low cost airline and (b) network airline only. Each of the three samples is balanced, i.e., they contain equal numbers of reviews recommending and not recommending an airline. This allows to properly perform the analyses and to evaluate the different machine learning classifiers. For each review, we take into account the review's text, the airline discussed and the overall recommendation decision.

3.3. Data preprocessing and projection

On airlinequality.com, travelers also have the ability to provide star ratings to express their evaluation about specific service aspects regarding their flight experience with the airline such as seat comfort and ground service. Nevertheless, as these fields are not mandatory, they are not provided in every review, and thus missing in a significant portion of the reviews in our dataset. As such, these ratings cannot be used reliably as variables or predictors of customer recommendation in further analysis. However, this again highlights the need for alternative mechanisms to derive recommendations related to service aspects that can be integrated in performance dashboards, and to further determine overall key performance indicators, for instance, in form of promoter scores. In the absence of specific star ratings corresponding to each service aspect, we perform an automated content analysis of the different online reviews' texts guided by the AD and MPAA theoretical models. As shown by previous research, the opinion expressed within online reviews can be extracted by means of sentiment analysis [53,

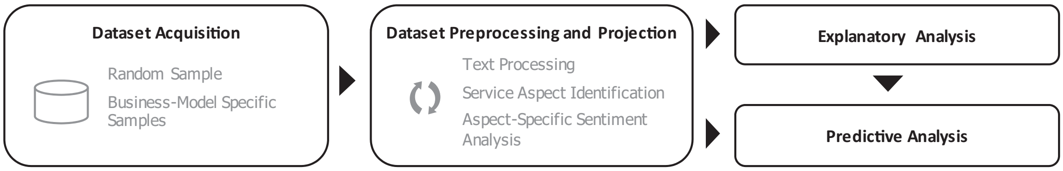
Dataset Acquisition
Random Sample
Business-Model Specific Samples
Dataset Preprocessing and Projection
Text Processing
Service Aspect Identification
Aspect-Specific Sentiment Analysis
Explanatory Analysis
Predictive Analysis

Fig. 1. Research process followed.

54]. We extend this approach further by aiming to extract the sentiment expressed towards the service as well as each of its specific aspects.

In order to extract the sentiment expressed within a review, we leverage the Harvard General Inquirer lexicon that connects syntactic, semantic, and practical information to words (e.g., delay is tagged having a negative sentiment) [55]. We specifically take into account the word lists for positive (pos) and negative (neg) words that are used to determine the sentiment polarity expressed within the different reviews, as shown in Eq. (1) [56]. We summarize the occurrences of positive and negative words while taking negations into consideration (in case of a negation preceding the sentiment-bearing term, its orientation is reversed). As shown in Eq. (1), sentiment polarity ranges from −1 (negative) to 1 (positive). Arguably, several terms in a review might indicate neutral sentiment. However, based on the AD theoretical model, we note that the diagnosticity, i.e., polarity of sentiment of a service-aspect, rather than a neutral stance, is likely to influence its recommendation by the reviewer.

$$\text{Polarity} = \frac{\text{pos}-\text{neg}}{\text{pos}+\text{neg}} \qquad (\text{Range}: -1 \text{ to } 1) \tag{1}$$

In order to further focus on the different service aspects within an online review, we additionally determine aspect-oriented sentiment measures. Given that such measures have not been previously applied in the field of airline reviews, we adopt a two-step approach: First, we *identify service aspects and develop domain-specific word lists* for topic detection in order to determine whether a specific part of a review deals with a certain topic (aspect) of interest. Next, we *determine the sentiment polarity for each service aspect* focusing on parts of the review that actually deal with those topics of interest. These two steps are described next.

For *identifying the service aspects expressed and for developing domain-specific word lists* that can be applied for topic detection, we adapt the approach proposed by Loughran and McDonald [57]. We analyze the different words which are contained in the whole corpus of online reviews. We manually analyze each word occurring in >2.5% of the online reviews. On the one hand, we check whether it belongs to the main airline service aspects that represent evaluation categories available for rating flight experiences on airlinequality.com, namely: *Seat Comfort*, *Cabin Staff Service*, *Inflight Entertainment*, *Food* & *Beverages*, *Ground Service*, *Value For Money*, *and Wifi* & *Connectivity* and map these to the core and augmented service categories, i.e. whether they relate to the core service (e.g. seat comfort) or whether they can be regarded as ancillary (e.g. food or entertainment, see Table 1). On the other hand, we focus on the core and augmented service concept as well as past empirical studies (i.e. Aksoy, Atilgan, and Akinci [58], Anderson, Pearo, and Widener [59], Chen and Chang [60] and Gilbert and Wong [61]) to identify aspects beyond the categories available on airlinequality.com. Following this procedure, we also identify *Punctuality*, *Safety*, *and Aircraft* as additional service aspects of interest. We recognize *Value for Money* as neither a core nor an augmented service aspect. Table 1 summarizes the resulting word lists.

After creating the word lists, we *determine the sentiment polarity for each service aspect*. For that purpose, we split a review into sentence-level units, and analyze whether at least one term related to a service aspect is contained in the sentences. For each sentence of the review fulfilling this condition, we then calculate the sentiment polarity according to Eq. (1). Finally, we determine the average sentiment polarity related to each aspect on a review-level. The specific variable operationalization is summarized in Table 2.

3.4. Explanatory analysis

In order to analyze the recommendation decision, we perform a logistic regression analysis as it is suitable for binary dependent variables. We analyze three different models, each explaining the binary dependent variable of whether a specific airline has been recommended by the reviewer or not. Model 1 (Eq. (2)) only takes into account the review's length, i.e., number of words in the review, as a control variable. It provides a baseline for comparing Models 2 and 3. Model 2 (Eq. (3)) takes into account the overall sentiment expressed within the review as an explanatory variable, while also controlling for the review's length. Finally, Model 3 (Eq. (4)) takes into account each of the aspect-oriented sentiment variables for explaining the decision to recommend a specific airline, also while controlling for the review's length. In all models, standard errors are clustered at airline level.

$$\begin{aligned}&\Pr(Recommended = 1) = F(constant + \beta_1 words + \varepsilon)\\&where\ F\left(\beta' X\right) = e^{\beta' X} / \left(1 + e^{\beta' X}\right)\end{aligned} \tag{2}$$

$$\begin{aligned}&\Pr(Recommended = 1) = F(constant + \beta_1 overall\_sentiment + \beta_2 words + \varepsilon)\\&where\ F\left(\beta' X\right) = e^{\beta' X} / \left(1 + e^{\beta' X}\right)\end{aligned} \tag{3}$$

Table 1
Developed word lists to extract service aspects from online reviews.

Category		Name	Identifying words
Core service aspects	Aircraft	Aircraft	Aeroplane, airplane, airbus, aircraft, Boeing, plane
	Seat comfort	Seat	Legroom, room, seat, space
	Safety	Safety	Positive: reliable, safe, stable negative: unreliable, unsafe, instable
	Punctuality	Punctuality	Positive: punctual, on time, quick, on schedule negative: delay, cancel, wait, miss, reschedule, late, postpone, slow
Augmented service aspects	Ground service	ground_service	Baggage, ground, lounge, terminal
	Cabin staff service	cabin_staff	Crew, staff
	Food & beverages	food_beverages	Beverage, breakfast, coffee, dinner, drink, food, lunch, meal, sandwich, snack
	Inflight entertainment	Entertainment	Entertainment, movie, TV
	Wifi & connectivity	wifi	Online, wifi, wi-fi
Value for money		value_money	Cost, pay, price, ticket, value

Table 2
Variable operationalization of independent (IV) and dependent variables (DV).

Type	Variable	Description	Definition
DV	Recommended	Binary variable measuring the reviewer's recommendation of an airline.	0 = not recommended 1 = recommended
IV	Overall_sentiment	Measures the overall sentiment polarity expressed within the online review, based on the positive and negative word lists of the General Inquirer.	(pos − neg) / (pos + neg) Range: −1, 1
	Aircraft Seat Safety Punctuality Ground_service Cabin_staff Food_beverages Entertainment Wifi value_money	Aspect-oriented sentiment variables that measure the sentiment expressed (sentiment polarity based on the General Inquirer) specifically related to the different aspects of the airline (measured at a sentence level). Covered aspects are aircraft type, seats, safety, staff, entertainment, food and beverages, ground service, value for money, wifi, and punctuality.	Average sentiment, ranging from −1 to 1, expressed in those sentences where the specific aspect is discussed.
	Words	Measures the quantity of words the online review consists of.	Number of words of the review

$$\begin{aligned}&\Pr(Recommended=1)=F(constant+\beta_1 aircraft+\beta_2 seat\\&\quad+\beta_3 safety+\beta_4 punctuality+\beta_5 ground\_service+\beta_6 cabin\_staff+\beta_7 food\_beverages\\&\quad+\beta_8 entertainment+\beta_9 wifi+\beta_{10} value\_money+\beta_{11} words+\varepsilon)\\&where\ F\left(\beta' X\right)=e^{\beta' X}/\left(1+e^{\beta' X}\right)\end{aligned} \tag{4}$$

3.5. Predictive analysis

In assessing the ability of a review's text contents to accurately predict a reviewer's recommendation to other travelers, we build different predictive models and evaluate their performance. Towards that end, we propose specific *model configurations* incorporating different attributes to predict the airline recommendation. Furthermore, we evaluate different *machine learning techniques*. Finally, we validate our results with a recommended *evaluation methodology* in the machine learning field to ensure that the results are realistic and not an artifact of model overfitting.

3.5.1. Model configuration

The different model configurations are summarized in Table 3, and differ with respect to predictors used. Model Configurations A and B take into account the different sentiment variables. Whereas Configuration A takes into account the overall sentiment of the airline review, Configuration B takes into account the more granular service aspect-oriented variables. Particularly, only those specific aspect-oriented sentiment variables that are found to have a significant influence on the airline recommendation are considered as predictors. In that regard, Configuration B builds on the results from the explanatory analysis.

Finally, Configuration C uses a classical text mining approach, bag-of-words, that is based on the different words of the text and thus has a more comprehensive feature set. This model configuration is later used to compare the results for models involving sentiment analysis. To be able to generate the term document matrix, we preprocess the reviews by means of tokenization, stop word filtering, stemming, n-gram generation and feature selection [36].

Table 3
Model configurations.

Configuration	Name	Description
A	Overall-sentiment	Model that takes into account overall_sentiment as input variable
B	Aspect-specific-sentiment	Model that takes into account the different significant aspect-oriented sentiment variables into account
C	Bag-of-words	Classical text mining approach based on a bag-of-words model, takes into account a review's words

3.5.2. Machine learning techniques

We evaluate the performance of different machine learning techniques in predicting the recommendation decision. Towards that goal, we perform supervised learning and learn from pre-labeled examples, i.e. the online reviews and the indication of the reviewer expressing the airline recommendation. In this context, we concentrate on *Naïve Bayes* (*NB*) as a rather simple learning algorithm as well as *Neural Network* (*NN*) and *Support Vector Machine* (*SVM*) representing more complex learning algorithms [62].

Naïve Bayes represents a simple machine learning technology relying on the Bayes theorem. Classifiers built upon the Bayes theorem are assumed to be naïve as they assume the independence of the different input variables. In Naïve Bayes classifiers, instances are classified based on the joint probabilities of their input variables. Although Naïve Bayes classifiers are rather simple and rely on potentially unrealistic assumptions, they have nevertheless been proven to generally perform well and in fact have the advantage of requiring low computational effort and thus being more time-efficient [63].

Neural Networks consist of a variety of (computational) neurons appearing in interconnected input, hidden, and output layers, and are intended to mimic the behavior of human neural networks. To achieve this behavior, weights are assigned to the connections between different neurons. Furthermore, each neuron has an activation function which is used to process the input of the neuron. The output neuron uses, as input, the weighted sum of outputs from neurons in the previous layer (or input variables in case of the initial input layer), and applies the activation function to the input [52]. When a neural network is trained, the weights of the different neurons are updated so that the overall neural network's output corresponds to the actual classification [64]. In this study, we apply a feed-forward neural network using backpropagation. As activation function, we use the most commonly adopted sigmoid function [65].

Support Vector Machine [66] represents another machine learning technique that is based upon the principle of finding the maximum margin hyperplane that maximizes the distances between instances of different classes [62]. As a linear separation of observations is not always possible, transformations are conducted by means of kernel functions that enable a separation of the observations according to their assigned classes. As shown by Hsu, Chang, and Lin [67], the Radial Basis Function (RBF) represents an appropriate kernel. We select the parameters of the kernel function by means of the grid-search heuristic proposed by [67].

3.5.3. Evaluation methodology

In order to evaluate the different predictive models, we perform stratified 10-fold cross-validation [68]. This evaluation procedure is advantageous as it avoids overfitting based on the notion that classifier training and classifier testing are performed on separate observations [69]. Furthermore, previous research has found that 10-fold stratified cross-validation performs best for evaluating models trained with real-world datasets such as the one in this study [68].

Within stratified 10-fold cross-validation, the whole sample is split into 10 different parts with equal class distributions. Next, nine parts are used for classifier training, whereas the remaining part is used for testing. This procedure is repeated by changing the different folds, so that each part is used nine times for training and one time for testing. After each iteration, the classification performance is determined by the number of correctly (true positives (TP), true negatives (TN)) as well as incorrectly classified (false positives (FP), false negatives (FN)) examples. These are depicted in an illustrative confusion matrix shown in Table 4.

At the end of 10-fold cross-validation, different performance measures can be calculated. These performance measures are presented in Eqs. (5)–(8) (displayed for class positive: recommended = 1).

$$Accuracy = \frac{TP + TN}{TP + FP + TN + FN} \tag{5}$$

$$Recall = \frac{TP}{TP + FN} \tag{6}$$

$$Precision = \frac{TP}{TP + FP} \tag{7}$$

$$F_1 = 2\cdot\frac{Precision\cdot Recall}{Precision + Recall} \tag{8}$$

To summarize, accuracy measures the total number of observations classified correctly (TP + TN), divided by the total number of observations (TP + FP + TN + FN). Precision measures how precise a classifier is, i.e. whether an example classified to belong to a specific class actually belongs to that class (TP / (TP + FP)) respectively (TN / (TN + FN)). Similarly, recall measures the percentage of many observations of a specific class are actually identified to belong to that class (TP / (TP + FN), TN / (TN + FP)). The F_1 measure aggregates precision and recall using their harmonic mean. This is worth noting, as often, precision and recall are related to each other and a high precision is accompanied by a low recall, and vice versa.

4. Empirical study

4.1. Descriptive statistics

Table 5 shows the descriptive statistics of the sample of airline reviews analyzed within this study. We observe that the overall sentiment as well as most aspect-oriented sentiment variables differ between (a) online reviews that recommend an airline and (b) online reviews that do not recommend an airline.

Interestingly, the overall sentiment is much more positive for recommending reviews than for non-recommending reviews. Regarding the different aspects of the service offered, the descriptive statistics show that the friendliness of the cabin staff is usually evaluated with much more positive sentiment than the ground service. It is interesting to note that the sentiment expressed regarding the "value for money" aspect is positive even for non-recommending reviews. Finally, the number of words related to non-recommending reviews is significantly larger than the number of words of recommending reviews. This provides an indication that non-recommending reviews might provide a stronger rationale for not recommending a certain airline based on flight experiences.

Table 4
Illustrative confusion matrix.

	Actual *recommended*	Actual *not recommended*
Predicted *recommended*	TP	FP
Predicted *not recommended*	FN	TN

Focusing on the distinction between LCC and FSNC airlines, we observe that especially seat comfort as core service aspect as well as entertainment and food as augmented service aspects are evaluated much more positive in case of FSNC Airlines. This is specifically related to the business model of LCC airlines, which provide fewer amenities for lower prices. Interestingly, the punctuality is slightly more positive in case of LCC than for FSNC which can also be explained by the business model of LCC as they try to implement more efficient business processes (e.g. shorter and more efficient ground handling) in order to increase profit.

Table 6 provides an overview on the correlations of the different variables taken into account. We note a high correlation between the overall sentiment of the review and the recommendation expressed by a reviewer. The correlations among the various service aspect-oriented sentiment variables are observed to be very low. Based on this, these variables can be assumed to be measuring distinct aspects of the review and do not indicate concern for multicollinearity when taken into account concurrently within the explanatory analysis. Furthermore, it can be observed that several aspect-oriented sentiment variables are moderately correlated with the overall sentiment expressed in the review. This observation validates the decision to consider the overall sentiment and the aspect-oriented variables in separate explanatory models (Eqs. (3) and (4)) to avoid potential side-effects.

4.2. Explanatory analysis

Table 7 shows the results of the explanatory analysis. In this analysis, we first take into account the control variables only (Model 1), overall sentiment (Model 2), as well as the aspect-specific sentiment in order to disentangle the recommendation decision, also focusing on the business model (Models 3 through 5). Furthermore, we also determine the explanatory power of core and augmented service aspects.

Focusing on Model 1 (Eq. (2)), which only takes into account the number of words of the review as a baseline setup, we observe that the number of words has a negative influence on the question of whether a specific airline is recommended (significant at a 1% level). Consequently, it can be argued that reviewers not recommending a specific product put more emphasis on describing the specific reasons of their decision, which in turn increases the review's length.

In case of Model 2 (Eq. (3)) that takes into account the overall sentiment expressed within the review, we observe a positive impact of the overall sentiment expressed on the reviewer's decision to recommend the airline service (significant at a 1% level). In other words, reviewers who are more positive about a flight experience are more likely to recommend the airline to others. This result shows that textual aspects of the reviews are valuable.

Model 3 (Eq. (4)) provides further insights into the drivers of the airline recommendation by analyzing whether the sentiment expressed towards individual service aspects of the service influences the airline recommendation. The results indicate that the reviewer's perception of the airline's seats, the punctuality, the ground service offered, the friendliness of the cabin staff, the quality of the food offered as well as the entertainment have a positive influence on the recommendation decision. Consequently, if reviewers express a positive sentiment towards these aspects, they are likely to recommend the airline. Interestingly, the perception of the wifi connectivity and of the value for money have no significant influence on the recommendation decision, suggesting that once a traveler has decided to select a specific airline for a

Table 5
Descriptive statistics.

Variable	Random sample				LCC airline random sample		FSNC airline random sample	
	Airline not recommended		Airline recommended					
	Mean	SD	Mean	SD	Mean	SD	Mean	SD
Overall_sentiment	−0.024	0.381	0.388	0.358	0.166	0.421	0.154	0.414
Aircraft	−0.017	0.394	0.082	0.392	0.038	0.369	0.032	0.390
Seat	−0.050	0.434	0.214	0.499	0.082	0.426	0.112	0.486
Safety	0.034	0.192	0.030	0.171	0.035	0.189	0.025	0.191
Punctuality	−0.141	0.623	0.225	0.640	0.063	0.699	0.042	0.668
Ground_service	−0.031	0.254	0.081	0.321	0.020	0.268	0.023	0.319
Cabin_staff	−0.013	0.448	0.324	0.500	0.153	0.471	0.143	0.479
Food_beverages	−0.042	0.459	0.264	0.484	0.064	0.332	0.131	0.482
Entertainment	0.030	0.313	0.151	0.381	0.020	0.204	0.088	0.353
Wifi	0.002	0.154	0.004	0.145	0.010	0.183	0.007	0.142
Value_money	0.048	0.399	0.088	0.365	0.087	0.397	0.069	0.358
Words	139.124	89.804	97.466	58.684	119.568	75.792	132.782	81.123

specific price level, other aspects of the service experience become more salient determining the perceived experience.

Taking into account the differences in case of LCC and FSNC airlines, we observe that customers focusing on LCC airlines do not focus on entertainment or ground service when forming their recommendation decisions. Instead, value for money has a slightly negative effect and safety is important. In case of FSNC airlines, the aircraft has a positive influence.

Regarding the overall quality of our results, the null hypothesis that none of the explanatory variables have an impact on the airline recommendation is rejected with a high level of confidence in all models ($p < .01$) in favor of the alternative hypothesis. Thus, results of the analysis can be regarded as valuable. Furthermore, in regards to Model 2, it can be noted that taking into account the overall sentiment considerably improves the model compared to the baseline (Regression 2, Δ Pseudo $R^2 = +0.168$). Furthermore, with Model 3, we find that taking into account the sentiment related to the different airline service aspects separately also improves the model compared to the baseline (Regression 3, Δ Pseudo $R^2 = +0.217$). As the increase in Pseudo R^2 for Model 3 is higher than in case of Model 2, this shows that the sentiments of service-specific aspects contained in the review are important for explaining the recommendation decision. Nevertheless, as also shown by the results, this increase in explanatory power is also dependent on the business model.

Finally, considering the explanatory power of core and augmented service aspects, we observe that augmented service aspects have a higher influence on the explanatory power than core service aspects. However, when analyzing based on the business model, we also find that while augmented service aspects are almost equally important in case of LCC and FSNC airlines (influence on Pseudo $R^2 + 0.091$ and $+0.087$), core service aspects are much more important in case of FSNC airlines (influence on Pseudo $R^2 + 0.054$ and $+0.080$).

4.3. Predictive analysis

The results of the predictive analysis in case of the random airline sample are shown in Table 8. The results show that the classifiers built upon the overall review sentiment (Classifiers A–C) as well as the individual service aspect-specific sentiment (Classifiers D–F) have a comparable predictive accuracy (up to 75.80%). Further, classifiers based on the bag-of-words features (Classifiers G–I) exhibit very good performance when predicting the overall recommendation hidden in unstructured online service reviews (predictive accuracy up to 80.80%). These results show that machine learning classifiers trained on the bag-of-words features as well as sentiment-based features allow to predict whether an online reviewer would ultimately recommend a specific airline and that the bag-of-words model outperforms the other configurations.

Tables 9 and 10 show the results of the classifier evaluation in case of the LCC and FSNC airline sample. Here, the sentiment-based classifiers show comparable performance. Interestingly, in case of the bag-of-words classifiers, a slightly improved performance (predictive accuracy up to 82.60% and 85.00% for LCC and FSNC samples respectively) can be observed. This can be attributed to the more specific adaptation to the specific language used in case of LCC and FSNC airline reviews.

Regarding the different machine learning algorithms, we observe that Neural Networks and SVM perform slightly better than Naïve Bayes in predicting the airline recommendation from unstructured user-generated online reviews. Interestingly, it is observed that classifiers taking into account only the overall sentiment achieve a

Table 6
Variable correlations.

	Variable	1	2	3	4	5	6	7	8	9	10	11	12	13
1	Recommended	1.00												
2	Overall_sentiment	0.49	1.00											
3	Aircraft	0.13	0.28	1.00										
4	Seat	0.27	0.43	0.19	1.00									
5	Safety	0.28	0.26	0.12	0.09	1.00								
6	Punctuality	−0.01	0.04	0.00	−0.07	0.03	1.00							
7	Ground_service	0.19	0.20	0.09	0.11	0.04	0.02	1.00						
8	Cabin_staff	0.33	0.45	0.12	0.16	0.17	0.06	0.17	1.00					
9	Food_beverages	0.31	0.41	0.11	0.21	0.04	−0.04	0.07	0.23	1.00				
10	Entertainment	0.17	0.24	0.04	0.12	0.03	−0.01	0.01	0.11	0.21	1.00			
11	Wifi	0.01	0.07	−0.01	0.06	0.04	0.00	0.06	0.00	0.06	0.01	1.00		
12	Value_money	0.05	0.24	0.09	0.12	0.04	−0.01	0.03	0.01	0.08	0.04	0.04	1.00	
13	Words	−0.27	−0.18	−0.07	−0.11	−0.04	0.09	−0.05	−0.07	−0.04	0.02	−0.02	0.04	1.00

Table 7
Logit regression results
Explaining airline recommendations by means of textual review aspects (n = 1000).

Variable	(1) Base model (Random sample)		(2) Overall sentiment (Random sample)		(3) Aspect-oriented sentiment (Random sample)		(4) Aspect-oriented sentiment (LCC sample)		(5) Aspect-oriented sentiment (FSNC sample)	
	Coef.	p-Value	Coef.	p-Value	Coef.	p-Value	Coef.	p-Value	Coef.	p-Value
Constant	0.937	<.01***	0.235	.157	0.567	<.01***	0.447	.06*	0.343	.49
Overall_sentiment			2.892	<.01***						
Aircraft					0.011	.96	0.515	.16	0.423	.01**
Seat					0.822	<.01***	0.466	<.01***	0.948	<.01***
Safety					−0.009	.98	0.715	<.01***	0.647	.16
Punctuality					0.860	<.01***	0.765	<.01***	0.949	<.01***
Ground_service					1.312	<.01***	0.349	.21	0.377	.23
Cabin_staff					1.052	<.01***	1.657	<.01***	1.025	<.01***
Food_beverages					1.164	<.01***	0.951	<.01***	1.124	<.01***
Entertainment					0.853	<.01***	0.195	.23	0.706	<.01***
Wifi					−0.627	.28	0.646	.15	0.857	.39
Value_money					0.108	.66	−0.377	.01**	0.106	.49
Words	−0.008	<.01***	−0.006	<.01***	−0.009	<.01***	−0.007	<.01***	−0.006	<.01***
$p > \chi^2$		<.01***		<.01***		<.01***		<.01***		<.01***
Pseudo R^2		.056		.224		.273		.221		.256
Δ Pseudo R^2				+.168		+.217				
Δ core						+.058		+.054		+.080
Δ augmented						+.117		+.091		+.087

Significance Levels: *** p < .01, ** p < .05, * p < .10.

comparable predictive accuracy when compared with classifiers focusing on aspect-oriented sentiment. Furthermore, from the point of view of predicting the consumer recommendation with high accuracy as well as with high precision and recall, it makes sense to focus on the bag-of-words classifier. Thus, classifiers built upon the contents of the online review are valuable to assess reviews that do not contain specific hints regarding the reviewer's recommendation decision. Such a predictive model can be used to compute promoter scores directly based on customer sentiments in online reviews.

Last, but not least, the transparency gained from using individual service aspect-oriented sentiment measures cannot be understated as well. In contrast to a bag-of-words-model, a model based on aspect-oriented sentiment breaks down the online review to specific metrics instead of taking into account the single words in the review. Thus, models using aspect-oriented sentiment metrics as performance indicators for specific service aspects are very applicable in enterprise performance dashboards as they provide more transparency although accompanied with slightly lower performance. The aforementioned approaches can thus be conceived to be on a spectrum with three levels: (a) highest predictive accuracy with little to no understanding of the underlying predictors, (b) relatively high predictive accuracy with high-level understanding of prediction based on overall sentiment expressed, (c) relatively high predictive accuracy with granular depiction of sentiment for various service aspects as contributors to the prediction.

5. Discussion

With this study, we first identify the different core and augmented service aspects expressed in online reviews focusing on airline services (referring to RQ1a). The results of the study clearly show that a reviewer's recommendation can be explained by service aspects expressed in online reviews (RQ1b) and thus confirm our theoretical reasoning based on the accessibility-diagnosticity as well as the multiple pathway anchoring and adjustment model. As shown by the explanatory analysis, the sentiment expressed within airline reviews has a significant influence on the question of whether a reviewer recommends a specific airline or not. Furthermore, the reviewer's perception of the specific aspects of the service offered, i.e. the sentiment expressed regarding aspects like food or ground service, also has a significant influence on the expressed recommendation. From a managerial standpoint, the service aspects that are found to be significant in influencing customer recommendation can form the basis for actionable improvements in the airline service offerings, and can guide further customer-based research studies.

In addition, as shown by the predictive analysis, we demonstrate that aspect-specific sentiment extracted from the reviews also has predictive power in case of forecasting airline recommendations (RQ2). We show that predictive models taking into account specific airline-related aspects perform well and that the generic bag-of-words model is valuable. Here, bag-of-words classifiers directly adapt to the language used

Table 8
Classifier Evaluation, Random Sample
Metrics are based on stratified 10-fold cross-validation.

				Recommended			Not recommended		
Classifier	Configuration	Algorithm	Accuracy	Prec.	Recall	F1	Prec.	Recall	F1
A	Overall-sentiment	Naïve Bayes	72.70	70.38	78.40	74.17	75.62	67.00	71.05
B		Neural network	75.00	73.23	78.80	75.91	77.06	71.20	74.01
C		SVM	74.70	74.26	75.60	74.92	75.15	73.80	74.47
D	Aspect-specific-sentiment	Naïve Bayes	73.90	75.59	70.60	73.01	72.42	77.20	74.73
E		Neural network	73.80	74.79	71.80	73.26	72.88	75.80	74.31
F		SVM	75.80	76.54	74.40	75.45	75.10	77.20	76.14
G	Bag-of-words	Naïve Bayes	77.70	75.51	82.00	78.62	80.31	73.40	76.70
H		Neural network	75.70	71.82	85.60	78.11	81.27	66.80	73.33
I		SVM	80.80	80.43	81.40	80.91	81.17	80.20	80.68

Table 9
Classifier evaluation, LCC airline sample
Metrics are based on stratified 10-fold cross-validation.

				Recommended			Not recommended		
Classifier	Configuration	Algorithm	Accuracy	Prec.	Recall	F1	Prec.	Recall	F1
A	Overall-sentiment	Naïve Bayes	71.30	71.43	71.00	71.21	71.17	71.60	71.38
B		Neural network	70.20	70.12	70.40	70.26	70.28	70.00	70.14
C		SVM	69.50	71.71	64.40	67.86	67.70	74.60	70.98
D	Aspect-specific-sentiment	Naïve Bayes	69.80	77.05	56.40	65.13	65.62	83.20	73.37
E		Neural network	73.20	76.24	67.40	71.55	70.79	79.00	74.67
F		SVM	73.70	77.12	67.40	71.93	71.05	80.00	75.26
G	Bag-of-words	Naïve Bayes	81.60	82.24	80.60	81.41	80.98	82.60	81.78
H		Neural network	80.00	85.89	71.80	78.22	75.77	88.20	81.51
I		SVM	82.60	83.27	81.60	82.43	81.96	83.60	82.77

for a specific category of airline and thus perform slightly better than sentiment-based classifiers. In contrast, sentiment-based classifiers can be considered to be more general as they focus on more general textual aspects. However, the primary advantage of these classifiers is the ease of understanding of their input variables in terms of the different service categories that has a clear pragmatic interpretation, compared to the different words of the text within the bag-of-words classifiers.

The classifiers built upon the overall sentiment perform comparable to classifiers based on the specific aspect-related sentiment scores. This shows that for predicting the recommendation decision with high-level sentiment understanding of the review, an overall analysis of the review may be sufficient. Nevertheless, specific scores are valuable if an airline wants to understand drivers of the consumer evaluation of different service aspects. The predictive analysis and the results of the stratified 10-fold cross-validation suggest that although the Pseudo R^2 of the explanatory analysis is not very high, a satisfactory classification performance can be achieved by applying machine learning methodologies.

In case of LCC and FSNC airlines (RQ3), slightly differing results are observed when focusing on the differences of the business model, mainly concerning the amenities offered. In that regard, core and augmented service aspects are found to be of relevance. The results suggest that LCC airlines should particularly improve augmented service aspects in order to foster consumer recommendations, which is interesting as the business model of LCC airlines is typically focused on providing the core service. Also, interestingly, value for money has no (or even a slightly negative) impact on the recommendation decision. A possible rationale for this result is the notion that as soon as a specific airline is selected according to a customer's willingness to pay, other aspects come to the forefront when deciding whether an airline can be recommended or not. If mentioned in the case of LCC airlines, 'value for money' can even have a negative impact. This can be explained with the observations made from a qualitative analysis of such reviews, where a consumer acknowledges the good value for money, but also criticizes other service aspects (e.g. "the prices are great, but…"). Additionally, the changing environmental conditions have now blurred the lines between LCC and FSNC to some extent, where LCC ticket prices are not always drastically cheaper than FSNCs, mainly due to operating efficiency gains by FSNCs over the years.

We are also aware of several limitations of our study. First, we are aware of the risk of overfitting related to predictive models which might lead to overoptimistic results. In order to avoid the results in this study being influenced by overfitting, we specifically take care that the predictive models are never trained and evaluated by taking into account the same observations. This is ensured by applying stratified 10-fold cross-validation model evaluation strategy that alleviates the risk of model overfitting.

Due to our supervised learning approach, the different online reviews under investigation have to include the recommendation decision ("yes" or "no") in order to be able to train and evaluate the model. Based on the 10-fold cross-validation evaluation, the results are generalizable out-of-sample and can be considered to also hold when the recommendation decision is not given at all – assuming that the structure of the online reviews is the same. Despite this issue, through our approach we demonstrate how to extract the specific service aspects from online reviews, which is valuable by itself, for instance, to aggregate the information given in numerous online reviews and to display this information in performance dashboards. Nevertheless, realistically, we assume this structural change to be unlikely, given that the main aim of online reviews, i.e., evaluating the product or service and providing support during the purchase decision making process, can be expected to be the same in all cases.

We are also cognizant that other factors apart from aspects discussed in the online review might influence the recommendation of a service. For instance, specific aspects of an airline such as the image of the airline's home country might also affect the recommendation. Given that such background information is not available, we indirectly cover this aspect by clustering the standard errors in the explanatory analysis taking into account the different airlines.

As previous research has shown, there might be a rating inflation regarding online review star ratings [70]. Assuming that online reviews constantly get more positive, it can also be presumed that this is resembled in the service aspects as well as in the question of whether a

Table 10
Classifier Evaluation, FSNC Airline Sample
Metrics are based on stratified 10-fold cross-validation.

				Recommended			Not recommended		
Classifier	Configuration	Algorithm	Accuracy	Prec.	Recall	F1	Prec.	Recall	F1
A	Overall-sentiment	Naïve Bayes	74.00	73.53	75.00	74.26	74.49	73.00	73.74
B		Neural network	72.70	72.65	72.80	72.72	72.75	72.60	72.67
C		SVM	75.00	74.51	76.00	75.25	75.51	74.00	74.75
D	Aspect-specific-sentiment	Naïve Bayes	74.60	76.39	71.20	73.70	73.03	78.00	75.43
E		Neural network	73.70	73.94	73.20	73.57	73.20	74.20	73.70
F		SVM	75.30	76.41	73.20	74.77	74.20	77.40	75.77
G	Bag-of-words	Naïve Bayes	82.30	78.99	88.00	83.25	86.46	76.60	81.23
H		Neural network	84.60	82.64	87.60	85.05	86.81	81.60	84.12
I		SVM	85.00	84.45	85.80	85.12	85.57	84.20	84.88

specific product or service is recommended, where the percentage of customers recommending the product or service might also be increased. Thus, as both variables can be assumed to increase, realistically, even in case of rating inflation, the observed relationships can be assumed to hold.

Finally, we recognize that our study analyzes a sample of airline reviews and thus, the proposed methodology of determining aspect-oriented sentiment is tailored to the aspects determining the decision to recommend a specific airline. Consequently, if the methodology is to be applied to services other than those offered by airlines, other aspects have to be taken into account, which also leads to the fact that additional word lists have to be developed for aspects that were not deemed relevant in the context of airline reviews. Nevertheless, as shown in this study, the proposed methodology to develop word lists related to specific service aspects leads to valuable measures of aspect-oriented sentiment and can thus also be followed within these contexts.

6. Conclusion

Within this study, we investigate whether user-generated content in form of online reviews can be leveraged to explain and predict the recommendation decision. Building upon the accessibility-diagnosticity model and the multiple pathway anchoring and adjustment model, we argue that the consumer recommendation of service is a collective expression of the sentiments regarding distinct service aspects, and emphasized by particularly those aspects that are perceived as salient during the service experience. Based on this rationale, we conduct explanatory and predictive analyses in order to analyze the drivers of the recommendation decision.

We find that the overall sentiment expressed in the review is significantly related to the question of whether a reviewer actually recommends a specific service. When disentangling the overall sentiment to the sentiment expressed towards the different service aspects, we show that aspect-related sentiment on core as well as augmented service aspects influences the recommendation decision as well and thus provides valuable information regarding the service dimensions. By means of a predictive analysis, we show that a bag-of-words model is best for predicting consumer recommendations with high accuracy. Furthermore, we show that the sentiment-related variables extracted from the review are valuable for providing transparent predictions of the recommendation decision. Finally, we also observe that the specific business model has an impact on explaining and predicting consumer recommendations.

Through this study, we contribute to the body of knowledge in several ways. First, we contribute to the literature explaining the recommendation decision by disentangling the recommendation of airlines: we show that the different core and augmented service-aspects offered by the airline are elaborated upon and evaluated in the service review and that the recommendation is driven by these aspects. We also contribute to the literature on service or product reviews which has so far mostly focused on the perceived helpfulness as well as on sales impact. With this study, we extend the previous understanding of such user generated content by focusing on the service recommendation expressed within the online review. Finally, we propose a novel approach to disentangle the overall sentiment expressed in the review and to take into account different aspects of a product or service.

The study allows stakeholders in the tourism ecosystem, particularly those related to airline services, to leverage the power of user generated content in the form of online reviews. In that regard, this study is highly relevant for practitioners. First, the predictive models proposed within this study can be used to classify online reviews without a specific indication of whether a service is recommended or not. Such a feature is very important for corporations that want to calculate a promoter score. Furthermore, this methodology enables corporations to take into account the opinions of a larger number of consumers than in case of directly contacting a panel of (potential) customers. The developed word lists can be used in order to automatically detect contents and related sentiments within online reviews. In that respect, our research allows suppliers of tourism related services to utilize machine learning algorithms to gain insights into user evaluations hidden in large amounts of unstructured, user-generated content. Additionally, the proposed methodology can also be used in order to provide decision support to consumers during their purchase process.

Finally, as shown by the explanatory analysis, sentiment related to different service aspects also significantly influences the recommendation decision. Consequently, the proposed sentiment indicators are also valuable and can be used in performance dashboards to show the specific customer evaluation of specific service aspects. This is also important having in mind the price, intangibility, emotional involvement and the risks associated with tourism related services [34] and thus satisfies the demand of travelers for risk reducing recommender systems [71].

This study provides several avenues for further research: As the current study is focused on the recommendations of airlines, other areas of the tourism industry can be taken into account to further disentangle the recommendation decision. Furthermore, as the specific sentiment indicators are also relevant on a stand-alone basis, a future design-oriented study could evaluate the specific configuration of an appropriate performance dashboard for both consumers and suppliers of tourism related services. Further research may also investigate alternative text analysis methods based on topic detection and clustering to improve upon the service aspect identification. Finally, future research might analyze the question of whether survey-based promoter scores and scores based on textual analysis change in tandem or whether one of these measures can be used as an early indicator for changes in consumer perceptions.

References

[1] F.F. Reichheld, The one number you need to grow, Harvard Business Review 81 (12) (2003) 46–55.

[2] B.A. Sparks, H.E. Perkins, R. Buckley, Online travel reviews as persuasive communication: the effects of content type, source, and certification logos on consumer behavior, Tourism Management 39 (2013) 1–9.

[3] Q. Ye, R. Law, B. Gu, W. Chen, The influence of user-generated content on traveler behavior: an empirical investigation on the effects of e-word-of-mouth to hotel online bookings, Computers in Human Behavior 27 (2) (2011) 634–639.

[4] H.T. Rhee, S.-B. Yang, How does hotel attribute importance vary among different travelers? An exploratory case study based on a conjoint analysis, Electronic Markets 25 (3) (2015) 211–226.

[5] M.D. Sotiriadis, C. van Zyl, Electronic word-of-mouth and online reviews in tourism services: the use of twitter by tourists, Electronic Commerce Research 13 (1) (2013) 103–124.

[6] N. Korfiatis, E. García-Bariocanal, S. Sánchez-Alonso, Evaluating content quality and helpfulness of online product reviews: the interplay of review helpfulness vs. review content, Electronic Commerce Research and Applications 11 (3) (2012) 205–217.

[7] K.K. Kuan, K.-L. Hui, P. Prasarnphanich, H.-Y. Lai, What makes a review voted? An empirical investigation of review voting in online review systems, Journal of the Association for Information Systems 16 (1) (2015) 48–71.

[8] S.M. Mudambi, D. Schuff, What makes a helpful online review? A study of customer reviews on amazon.com, MIS Quarterly 34 (1) (2010) 185–200.

[9] M. Siering, J. Muntermann, What drives the helpfulness of online product reviews? From stars to facts and emotions, Proc. of the 11th International Conference on Wirtschaftsinformatik, Leipzig, Germany, 2013.

[10] J. Ozment, E.A. Morash, The augmented service offering for perceived and actual service quality, Journal of the Academy of Marketing Science 22 (4) (1994) 352–363.

[11] A. Ravald, C. Grönroos, The value concept and relationship marketing, European Journal of Marketing 30 (2) (1996) 19–30.

[12] N.A. Morgan, L.L. Rego, The value of different customer satisfaction and loyalty metrics in predicting business performance, Marketing Science 25 (5) (2006) 426–439.

[13] P. Marsden, A. Samson, N. Upton, Advocacy drives growth, Brand Strategy 198 (2005) 45–47.

[14] E. de Haan, P.C. Verhoef, T. Wiesel, The predictive ability of different customer feedback metrics for retention, International Journal of Research in Marketing 32 (2) (2015) 195–206.

[15] S. Khan, How Philips Uses Net Promoter Scores to Understand Customers, Harvard Business Review Online, 2011https://hbr.org/2011/05/how-philips-uses-net-promoter, Accessed date: 28 May 2016.

[16] J. He, H. Liu, H. Xiong, SocoTraveler: travel-package recommendations leveraging social influence of different relationship types, Information Management 53 (8) (2016) 934–950.

[17] A. De Bruyn, J.C. Liechty, E. Huizingh, G.L. Lilien, Offering online recommendations with minimum customer input through conjoint-based decision aids, Marketing Science 27 (3) (2008) 443–460.
[18] W. Wang, I. Benbasat, Attributions of trust in decision support technologies: a study of recommendation agents for E-commerce, Journal of Management Information Systems 24 (4) (2008) 249–273.
[19] B. Xiao, I. Benbasat, E-commerce product recommendation agents: use, characteristics, and impact, MIS Quarterly 31 (1) (2007) 137–209.
[20] S. Jang, A. Prasad, B. Ratchford, How consumers use product reviews in the purchase decision process, Marketing Letters 23 (3) (2012) 825–838.
[21] V. Dorner, O. Ivanova, M. Scholz, Think twice before you buy! How recommendations affect three-stage purchase decision processes, ICIS 2013 Proceedings, 2013.
[22] J.A. Chevalier, D. Mayzlin, The effect of word of mouth on sales: online book reviews, Journal of Marketing Research 43 (3) (2006) 345–354.
[23] C. Forman, A. Ghose, B. Wiesenfeld, Examining the relationship between reviews and sales: the role of reviewer identity disclosure in electronic markets, Information Systems Research 19 (3) (2008) 291–313.
[24] A. Ghose, P.G. Ipeirotis, Estimating the helpfulness and economic impact of product reviews: mining text and reviewer characteristics, IEEE Transactions on Knowledge and Data Engineering 23 (10) (2011) 1498–1512.
[25] T.L. Ngo-Ye, A.P. Sinha, The influence of reviewer engagement characteristics on online review helpfulness: a text regression model, Decision Support Systems 61 (2014) 47–58.
[26] R.M. Schindler, B. Bickart, Perceived helpfulness of online consumer reviews: the role of message content and style, Journal of Consumer Behaviour 11 (3) (2012) 234–243.
[27] D. Yin, S.D. Bond, H. Zhang, Anxious or angry? Effects of discrete emotions on the perceived helpfulness of online reviews, MIS Quarterly 38 (2) (2014) 539–560.
[28] L. Huang, C.-H. Tan, W. Ke, K.-K. Wei, Do we order product review information display? How? Information Management 51 (7) (2014) 883–894.
[29] C. Luo, X. Luo, Y. Xu, M. Warkentin, C.L. Sia, Examining the moderating role of sense of membership in online review evaluations, Information Management 52 (3) (2015) 305–316.
[30] J. Qi, Z. Zhang, S. Jeon, Y. Zhou, Mining customer requirements from online reviews: a product improvement perspective, Information Management 53 (8) (2016) 951–963.
[31] T. Hennig-Thurau, K.P. Gwinner, G. Walsh, D.D. Gremler, Electronic word-of-mouth via consumer-opinion platforms: what motivates consumers to articulate themselves on the Internet? Journal of Interactive Marketing 18 (1) (2004) 38–52.
[32] M. Siering, J. Muntermann, How to Identify Tomorrow's most active social commerce contributors? Inviting Starlets to the reviewer hall of fame, Proceedings of the 34th International Conference on Information Systems, Milan, Italy, 2013.
[33] N. Chung, H. Han, C. Koo, Adoption of travel information in user-generated content on social media: the moderating effect of social presence, Behaviour & Information Technology 34 (9) (2015) 902–919.
[34] R. Tilly, K. Fischbach, D. Schoder, Mineable or messy? Assessing the quality of macro-level tourism information derived from social media, Electronic Markets 25 (3) (2015) 227–241.
[35] Z. Xiang, Z. Schwartz, J.H. Gerdes, M. Uysal, What can big data and text analytics tell us about hotel guest experience and satisfaction? International Journal of Hospitality Management 44 (2015) 120–130.
[36] G. Miner, Practical Text Mining and Statistical Analysis for Non-structured Text Data Applications, Academic Press, 2012.
[37] A. Hotho, A. Nürnberger, G. Paaß, A brief survey of text mining, GLDV Journal of Computational Linguistics 20 (1) (2005) 19–62.
[38] S.S. Groth, M. Siering, P. Gomber, How to enable automated trading engines to cope with news-related liquidity shocks? Extracting signals from unstructured data, Decision Support Systems 62 (2014) 32–42.
[39] D. Lewis, Representation and Learning in Information Retrieval: Dissertation, University of Massachusetts, 1992.
[40] R. Feldman, J. Sanger, The Text Mining Handbook: Advanced Approaches in Analyzing Unstructured Data, Cambridge University Press, 2007.
[41] M. Hu, B. Liu, Mining and summarizing customer reviews, Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD-2004), Seattle, Washington, USA, 2004.
[42] A.-M. Popescu, O. Etzioni, Extracting product features and opinions from reviews, Natural language processing and text mining (2007) 9–28.
[43] H.-J. Jeong, D.-M. Koo, Combined effects of valence and attributes of e-WOM on consumer judgment for message and product: the moderating effect of brand community type, Internet Research 25 (1) (2015) 2–29.
[44] Y. Kwark, J. Chen, S. Raghunathan, Online product reviews: implications for retailers and competing manufacturers, Information Systems Research 25 (1) (2014) 93–110.
[45] J. Feldman, J. Lynch, Self-generated validity and other effects of measurement on belief, attitude, intention, and behavior, The Journal of Applied Psychology 73 (3) (1988) 421–435.
[46] Lynch Jr., G. John, H. Marmorstein, M.F. Weigold, Choices from sets including remembered brands: use of recalled attributes and prior overall evaluations, Journal of Consumer Research (1988) 169–184.
[47] J.B. Cohen, A. Reed, A multiple pathway anchoring and adjustment (MPAA) model of attitude generation and recruitment, Journal of Consumer Research 33 (1) (2006) 1–15.
[48] J.G. Lynch, Accessibility-diagnosticity and the multiple pathway anchoring and adjustment model, Journal of Consumer Research 33 (1) (2006) 25–27.
[49] Boeing, Airline Strategies and Business ModelsAvailable online http://www.boeing.com/commercial/market/long-term-market/airline-strategies-and-business-models/ 2016, Accessed date: 26 September 2016.
[50] D. Gillen, A. Lall, Competitive advantage of low-cost carriers: some implications for airports, Journal of Air Transport Management 10 (1) (2004) 41–50.
[51] J.F. O'Connell, G. Williams, Passengers' perceptions of low cost airlines and full service carriers: a case study involving Ryanair, Aer Lingus, Air Asia and Malaysia Airlines, Journal of Air Transport Management 11 (4) (2005) 259–272.
[52] J. Han, M. Kamber, Data Mining: Concepts and Techniques, 2nd ed. Elsevier; Morgan Kaufmann, San Francisco, 2006.
[53] R. Feldman, Techniques and applications for sentiment analysis, Communications of the ACM 56 (4) (2013) 82–89.
[54] B. Pang, L. Lee, Opinion mining and sentiment analysis, Foundations and Trends in Information Retrieval 2 (1–2) (2008) 1–135.
[55] P.J. Stone, D.C. Dunphy, M.S. Smith, The General Inquirer: A Computer Approach to Content Analysis, MIT Press, Cambridge, MA, 1966.
[56] P.C. Tetlock, M. Saar-Tsechansky, S. Macskassy, More than words: quantifying language to measure firms' fundamentals, The Journal of Finance 63 (3) (2008) 1437–1467.
[57] T. Loughran, B. McDonald, When is a liability not a liability? Textual analysis, dictionaries, and 10-Ks, The Journal of Finance 66 (1) (2011) 35–65.
[58] S. Aksoy, E. Atilgan, S. Akinci, Airline services marketing by domestic and foreign firms: differences from the customers' viewpoint, Journal of Air Transport Management 9 (6) (2003) 343–351.
[59] S. Anderson, L.K. Pearo, S.K. Widener, Drivers of service satisfaction linking customer satisfaction to the service concept and customer characteristics, Journal of Service Research 10 (4) (2008) 365–381.
[60] F.-Y. Chen, Y.-H. Chang, Examining airline service quality from a process perspective, Journal of Air Transport Management 11 (2) (2005) 79–87.
[61] D. Gilbert, R.K.C. Wong, Passenger expectations and airline services: a Hong Kong based study, Tourism Management 24 (5) (2003) 519–532.
[62] S.B. Kotsiantis, Supervised machine learning: a review of classification techniques, Informatica 31 (3) (2007) 249–268.
[63] P. Langley, W. Iba, K. Thompson, An analysis of Bayesian classifiers, Proceedings of the Tenth National Conference on Artificial Intelligence, Seattle, WA, 1992.
[64] R. Nisbet, J.F. Elder, G. Miner, Handbook of Statistical Analysis and Data Mining Applications, Academic Press/Elsevier, Amsterdam, Boston, 2009.
[65] C.M. Fuller, D.P. Biros, R.L. Wilson, Decision support for determining veracity via linguistic-based cues, Wireless Healthcare 46 (3) (2009) 695–703.
[66] V. Vapnik, The Nature of Statistical Learning Theory, Springer, New York, USA, 1995.
[67] C.W. Hsu, C.C. Chang, C.J. Lin, A Practical Guide to Support Vector Classification, National Taiwan University, 2003http://www.csie.ntu.edu.tw/~cjlin/papers/guide/guide.pdf, Accessed date: 16 October 2011.
[68] R. Kohavi, A study of cross-validation and bootstrap for accuracy estimation and model selection, Proceedings of the International Joint Conference on Artificial Intelligence, Montreal, Quebec, Canada, Vol. 14(2), , 1995.
[69] T. Mitchell, Machine Learning, McGraw-Hill, London, 1997.
[70] E. Wolff-Mann, When Did a 4-star Review Become a Bad Review? Time, 2016http://time.com/money/page/online-reviews-trust-fix/, Accessed date: 15 August 2017.
[71] D. Buhalis, A. Amaranggana, Smart tourism destinations enhancing tourism experience through personalisation of services, Information and Communication Technologies in Tourism 2015 (2015) 377–389.

Michael Siering is a postdoctoral research associate at Goethe University Frankfurt and a research associate at the E-Finance Lab, an industry-academic partnership between Goethe University Frankfurt and several industry partners. He has been a visiting scholar at Penn State University. His research focuses on decision support systems in electronic markets, with a focus on the analysis of user generated content by means of sentiment analysis and text mining. His work has been published in the Journal of Management Information Systems, Journal of Information Technology, Decision Support Systems and conference proceedings such as ICIS, ECIS and HICSS. He holds a M.Sc. and a Ph.D. in business administration from Goethe University Frankfurt.

Amit V. Deokar is an Assistant Professor of Management Information Systems in the Robert J. Manning School of Business at the University of Massachusetts Lowell. Dr. Deokar received his PhD in Management Information Systems from the University of Arizona. He also earned a MS in Industrial Engineering from the University of Arizona and a BE in Mechanical Engineering from VJTI, University of Mumbai. His research interests include data analytics, enterprise data management, business intelligence, business process management, and collaboration processes. His work has been published in journals such as Journal of Management Information Systems, Decision Support Systems (DSS), The DATA BASE for Advances in Information Systems, Information Systems Frontiers, Business Process Management Journal (BPMJ) and IEEE Transactions. He is currently a member of the editorial board of DSS and BPMJ journals. He has been serving as the Decision Support and Analytics Track Chair at the international AMCIS 2014–17 conferences, and is currently the Chair of the AIS Special Interest Group on Decision Support and Analytics (SIGDSA). He was recognized with the 2014 IBM Faculty Award for his research and teaching in the areas of analytics and big data.

Christian Janze is a doctoral candidate at Goethe University Frankfurt and a research assistant at the E-Finance Lab, an industry-academic partnership between Goethe University Frankfurt and several industry partners. In addition, he is part of the doctoral program of DZ BANK. He holds a Master's degree in Management from Goethe University Frankfurt and was a scholar of the German National Academic Foundation. His research focuses on user generated content, online reviews and market efficiency and has appeared in ICIS Proceedings.