

Article

Topic Modeling and Sentiment Analysis of Online Review for Airlines

Hye-Jin Kwon ¹, Hyun-Jeong Ban ², Jae-Kyoon Jun ¹ and Hak-Seon Kim ^{2,*} 

¹ Division of Business Administration, Pukyong National University, Busan 48513, Korea; dns207@naver.com (H.-J.K.); jkjun@pknu.ac.kr (J.-K.J.)

² School of Hospitality & Tourism Management, Kyungsoo University, Busan 48434, Korea; helenaban@ks.ac.kr

* Correspondence: kims@ks.ac.kr

Abstract: The purpose of this study is to conduct topic modeling and sentiment analysis on the posts of Skytrax (airlinequality.com), where there are many interests and participation of the people who have used or are willing to use it for airlines. The purpose of people gathering at Skytrax is to make better choices using the actual experiences of other customers who have experienced airlines. Online reviews written by customers with experience using airlines in Asia were collected. The data collected were online reviews from 27 airlines, with more than 14,000 reviews. Topic modeling and sentiment analysis were used with the collected data to figure out what kinds of important words are in the online reviews. As a result of the topic modeling, ‘seat’, ‘service’, and ‘meal’ were significant issues in the flight through frequency analysis. Additionally, the result revealed that delay was the main issue, which can affect customer dissatisfaction while ‘staff service’ can make customers satisfied through sentiment analysis as the result shows the ‘staff service’ with meal and food in the topic modeling.

Keywords: Asian airlines; online review; sentiment analysis; topic modeling; text mining; big data



Citation: Kwon, H.-J.; Ban, H.-J.; Jun, J.-K.; Kim, H.-S. Topic Modeling and Sentiment Analysis of Online Review for Airlines. *Information* **2021**, *12*, 78. <https://doi.org/10.3390/info12020078>

Academic Editor: Davide Buscaldi
Received: 11 January 2021
Accepted: 9 February 2021
Published: 12 February 2021

Publisher’s Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

The global airline industry is facing increased competition between airlines and regions due to the expansion of the Treaty on Open Skies, private participation in airline and airport operations, and the interaircraft partnerships and mergers [1]. To survive this competition, airlines continue to make efforts to improve service quality as survival strategies. Just as the product quality revolution in manufacturing determines a company’s competitiveness, in the service sector, service quality innovation is a factor that determines a company’s win or loss [2]. In addition, the development of service quality is perceived as a means of securing a competitive advantage with loyal customers. However, the airline industry is not aware of the customer’s needs, and the provision of quality of service is compromised. As a result, customer needs have consistently attracted the attention from scholars as a fundamental variable in customer service delivery, and should be more important to airlines to identify customer needs and provide the right quality of service [3,4].

The development of Internet technology is changing the hospitality industry, which puts customer needs first. Using the Internet, customers can communicate with businesses as well as deliver information to one another [5]. In particular, online reviews are having an important impact on the decision-making of potential purchasing customers rather than on corporate marketing activities, because online reviews left by experienced customers are recognized as objective and reliable information [6]. Therefore, various hospitality industries are using online reviews as a marketing tool. In addition, the hospitality industry has been able to use the Internet to easily access customer opinions and establish relationships

with customers [7]. These changes are also bringing significant alterations to the airline industry, which puts customer needs first.

The survey method has the advantage of being able to get answers to the questions you ask. However, there are limitations such as measurement errors that can occur in survey patterns, survey terminology, response categories, and survey order [8]. Additionally, there were many quantitative studies of structured survey methods in the research methodology. However, with limitations and problems regarding on existing quantitative survey methods, changes in research methodology are required [9]. In particular, due to the nature of the airline industry, sentiment analysis and topic modeling techniques are drawing attention as a new research method to understand consumers' in-depth thoughts as online reviews and comments from customers have a lot of influence on their purchasing intentions [10]. Topic modeling is an analysis that has been actively used in recent research trends analyses to deduce topics latent in text data through observed words and derive hidden meanings to understand the overall trend. Sentiment analysis, a branch of text mining, is called opinion mining as a natural language processing analysis method that identifies the opinions of positive or negative people posted in text data [11].

Recently, there have been various studies using sentiment analysis in general products, politics, and society, including the service industry, such as movies, tourism, restaurants, and hotels, where the importance of online reviews is emphasized [12–18]. Accordingly, the airline-related research also emphasizes the importance of big data analysis, and although sentiment analysis of customers' online reviews is under way, it is still only in the beginning stage. Therefore, this study conducted a sentiment analysis to identify the meaning of positive and negative reviews recorded online by customers using airlines in Asia. The purpose is to derive meaningful implications by analyzing what customers feel satisfied and dissatisfied with. It also seeks to provide an opportunity to assist airlines in their management activities and decision-making process, and to provide key data and analysis methods that are useful for their management.

This paper is expanded as follows; after the introduction, the subsequent literature review presents the previous works of Asian Airlines, online customer review, topic modeling, and sentiment analysis using big data. The methodology section explains the data collection and data analysis with big data analytics. The results of this paper are presented divided into frequency analysis, word cloud, topic modeling, and sentiment analysis. The final section summarizes the study and the implications of this study both for academia and practice; the limitations and future research directions are well indicated and discussed as well.

2. Literature Review

2.1. Asian Airline

Asia-Pacific airlines are an important collective force in the international aviation market, accounting for a quarter of the total global air passenger demand and two-fifths of the global air cargo demand [19]. Boeing predicts that Asia will account for half of global aviation demand growth [20]. According to IATA [21], in terms of profitability, Asia-Pacific airlines have achieved half the profits of the global aviation industry. The year 2010 accounted for about USD 10 billion out of USD 18 billion, and USD 2.1 billion out of USD 4 billion in 2011, when rising oil prices seriously weighed down the industry's profitability. Asian airlines, which have growing global influence and importance as a group, are expected to play an active role in creating the future global air transport industry.

The Asia-Pacific region is already the world's largest aviation market. Existing airlines in Asia are creating many new airlines to reach different segment markets. Part of this attempt is in the form of a joint venture between traditional airlines and new airlines, which combines their respective influences to gain new access to the international market. While Asia-Pacific aviation demand continues to grow, fierce competition is stifling airline revenue generation as the supply of airlines overflows into the market [22]. As Asia-Pacific airlines increase their supply, this is a very challenging business environment that puts downward

pressure on prices and profits. As a result, the import growth rates of many airlines in the region are flat and profitability is hard to find. In order to present marketing implications for airlines to survive in the highly competitive Asian aviation market, the research was conducted to effectively identify customer satisfaction and dissatisfaction and present marketing implications.

2.2. Online Review

Online reviews are called electronic word of mouth (eWOM), and customers who have experienced services freely describe or choose scores, which are perceived as more reliable and objective than information provided by companies. This is why online reviews play a major role in creating images before experiencing service to customers [23,24]. Unlike general reviews, the features of online reviews are accessible 24 h a day, enabling continuous information storage in text or images [25]. Not only has the coverage become broader, but the spread of information is fast. This means that online reviews have a significant impact on service-enabled customers because they can be written and edited without time and space constraints [26]. Through online reviews, customers also share reviews of their services with others through the online community regardless of their commercial interests. Unlike one-sided advertising, it also controls customers' choices because it conveys the customer's true opinions or specific experiences they experience [27]. Various information exchanges on airline services are also taking place actively through online reviews, as airline services are difficult to review before experience and are not easy to evaluate on an objective basis, as well as making careful choices with relatively high-priced services [28]. This has led to increased information acquisition and purchase of tickets over the Internet, and airlines have recently been building promotions or marketing strategies using mobile phones to compete with other airlines [29].

Many scholars realized the importance of online reviews and did research using online reviews. Dellarocas et al. [30] have demonstrated that the metrics of online review can accurately predict movie revenue. Sotiriadis and van Zyl [31] found that online reviews and recommendations affect the decision making process of tourists towards tourism services and WOM has a significant impact on the subjective norms and attitudes towards an airline, and a customer's willingness to recommend. M. Siering et al. [32] have investigated whether user-generated content in form of online reviews can be leveraged to explain and predict the recommendation decision. Additionally, they discovered sentiment related to different service aspects also significantly influences the recommendation decision. Gutierrez and Alsharif [33] investigated the tweets mining approach to detection of critical events. Therefore, the online review would be very useful for airlines to understand their diverse customers in order to take service improvement strategies since airlines are a highly competitive industry.

2.3. Topic Modeling and Sentiment Analysis Using Big Data

Latent Dirichlet allocation (LDA) is the most popular topic model, which is a method for analyzing a large set of documents. The basic idea is that documents are represented as a topic distribution where each topic is characterized by a word distribution. $p(z | d_i)$ is the topic probability density function for document i , and $p(w | z_{i,j})$ is the word probability density function for the topic assigned to the j th word in document i . Given this distribution, LDA creates a new document through the following generation process:

for j th word in the i th document:

Choose a topic $z_{i,j} \sim \text{Multinomial}(p(z | d_i))$

Choose a topic $w_{i,j} \sim \text{Multinomial}(p(w | z_{i,j}))$

Depending on what you read from the text, text data analysis can be largely divided into two categories: topic modeling and sentiment analysis. The topic modeling analysis refers to a series of techniques that identify what text deals with, and the sentiment analysis refers to a series of techniques that identify the emotions or sentiments that appear in the text [34]. Topic modeling is a technique that extracts and suggests potentially meaningful

topics from a great number of documents based on a procedural probability distribution model [35]. A great number of studies were conducted on various types of unstructured text documents, including SNS and online reviews, through topic modeling. Table 1 lists studies using topic modeling.

Table 1. Recent studies using topic modeling analysis.

Author	Year	Title	Main Point
Lucini, F.R.; Tonetto, L.M.; Fogliatto, F.S.; Anzanello, M.J. [36]	2020	Text mining approach to explore dimensions of airline customer satisfaction using online customer reviews	55,000 reviews covering 400 airlines and passengers from 170 countries analyzed using latent Dirichlet allocation (LDA) model, and identified 27 dimensions of satisfaction.
Kim, S.; Park, H.; Lee, J. [37]	2020	Word2vec-based latent semantic analysis (W2V-LSA) for topic modeling: A study on blockchain technology trend analysis	This paper applies LDA topic modeling to a vast number of passenger's online review to compare service quality between full service carriers and low cost carriers.
Sutherland, I.; Sim, Y.; Lee, S.K.; Byun, J.; Kiatkawsin, K. [38]	2020	Topic Modeling of Online Accommodation Reviews via Latent Dirichlet Allocation	This paper applied an inductive approach by utilizing large unstructured text data of 104,161 online reviews of Korean accommodation customers to frame which topics of interest guests find important.
Lim, J.; Lee, H.C. [39]	2019	Comparisons of service quality perceptions between full service carriers and low cost carriers in airline travel	This paper proposed a new topic modeling method called Word2vec-based Latent Semantic Analysis to perform an annual trend analysis of blockchain research by country and time for 231 abstracts of blockchain-related papers published over the past five years.
Sun, L.; Yin, Y. [40]	2017	Discovering themes and trends in transportation research using topic modeling	This paper applied a LDA model on article abstracts to infer 50 key topics. We show that those characterized topics are both representative and meaningful, mostly corresponding to established subfields in transportation research.

Sentiment analysis is also called opinion mining as one of the text mining analyses that extracts consumer emotions, opinions, attitudes, etc. Due to the recent development of Internet media such as SNS, e-commerce, and online communities, text data containing subjective elements is flooding online [41]. As a result, the importance of emotional analysis was highlighted as the user's sensibility extracted from text data was actively utilized in the enterprise's marketing [42,43]. Sentiment analysis establishes a sentiment dictionary consisting of emotional words and polarities indicating the degree of positivity and negation of words, and quantifies emotions using these sentiment dictionaries. A set of words, such as a sentiment dictionary, is very important to derive accurate sentiment analysis results. Liu [11] built Opinion Lexicon, which consists of English, to perform sentiment analysis by extracting 2006 positive and 4783 negative words over the years. In addition, Wibe et al. [44] created an MPQA sentiment dictionary that delicately defines emotion and sensitivity according to the purpose of the emotional vocabulary appearing in about 10,000 sentences. Esuli and Sebastiani [45], based on the existing set of WordNet synonyms, developed the SentiWordNet by distinguishing between three levels of sensitivity: positive, neutral, and negative as a result of the classification of semi-supervised.

3. Materials and Methods

Online review data posted on Skytrax (airlinequality.com) [46], the world's largest airport and airline service assessment site, was collected to provide an empirical analysis of this study. This work explores the latent meaning through the results of various text mining

techniques, focusing on online reviews left by experienced flying customers on the Skytrax. This has released well-known online reviews of the airline industry in terms of reliability and recognition to ensure objectivity in evaluating airline service quality. An annual Airline Customer Satisfaction Survey was used by the Skytrax to select targets for data collection. Among the World's Top 100 Airlines selected through the Customer Satisfaction Survey, airlines in Asia were selected for the study of airlines based on the airline.

The collected data was analyzed using the R program, which is an open source program. The procedures performed in topic modeling consist largely of three stages. First, collect the data to which topic modeling will be applied. In the second step, preprocessing and morphological analysis was performed to transform unstructured data collected into data suitable for topical modeling. The last step is data analysis. Frequency analysis of words derived from morphological analysis was conducted and word cloud was visualized [33]. Topic modeling and sentiment analysis were also performed by converting unstructured text data into a structured form, Document-Term Matrix (DTM). The detailed research procedures were carried out as shown in Figure 1 with three systems: data collection, text mining, and data analysis.

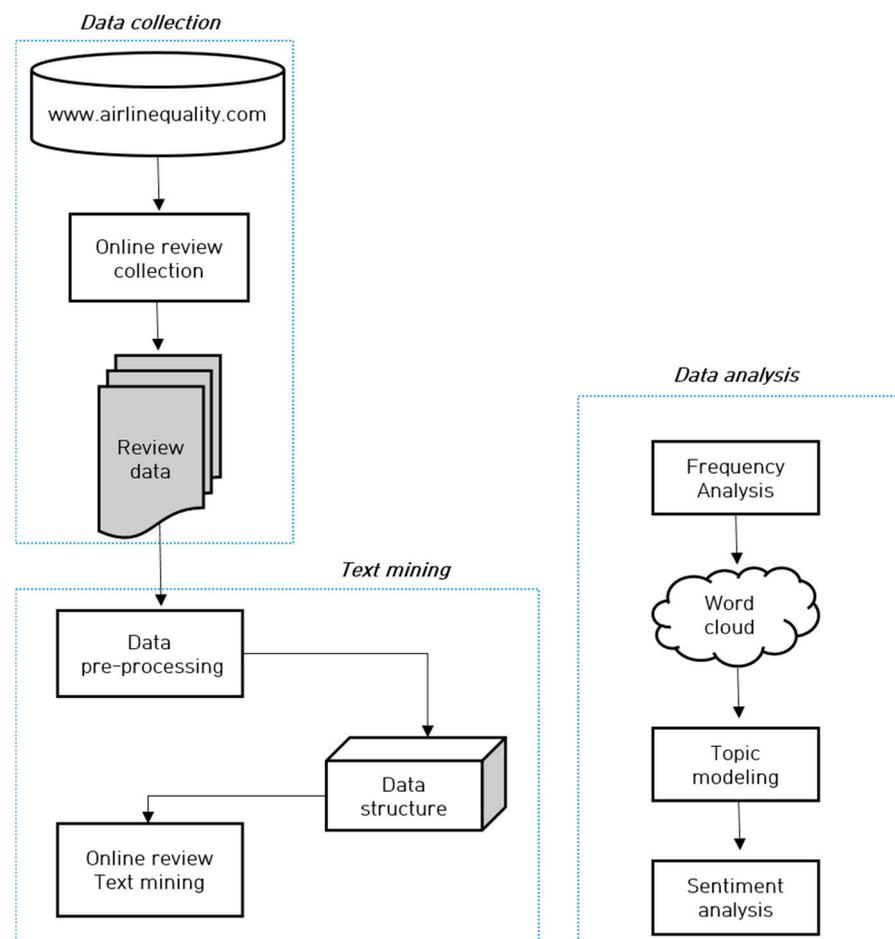
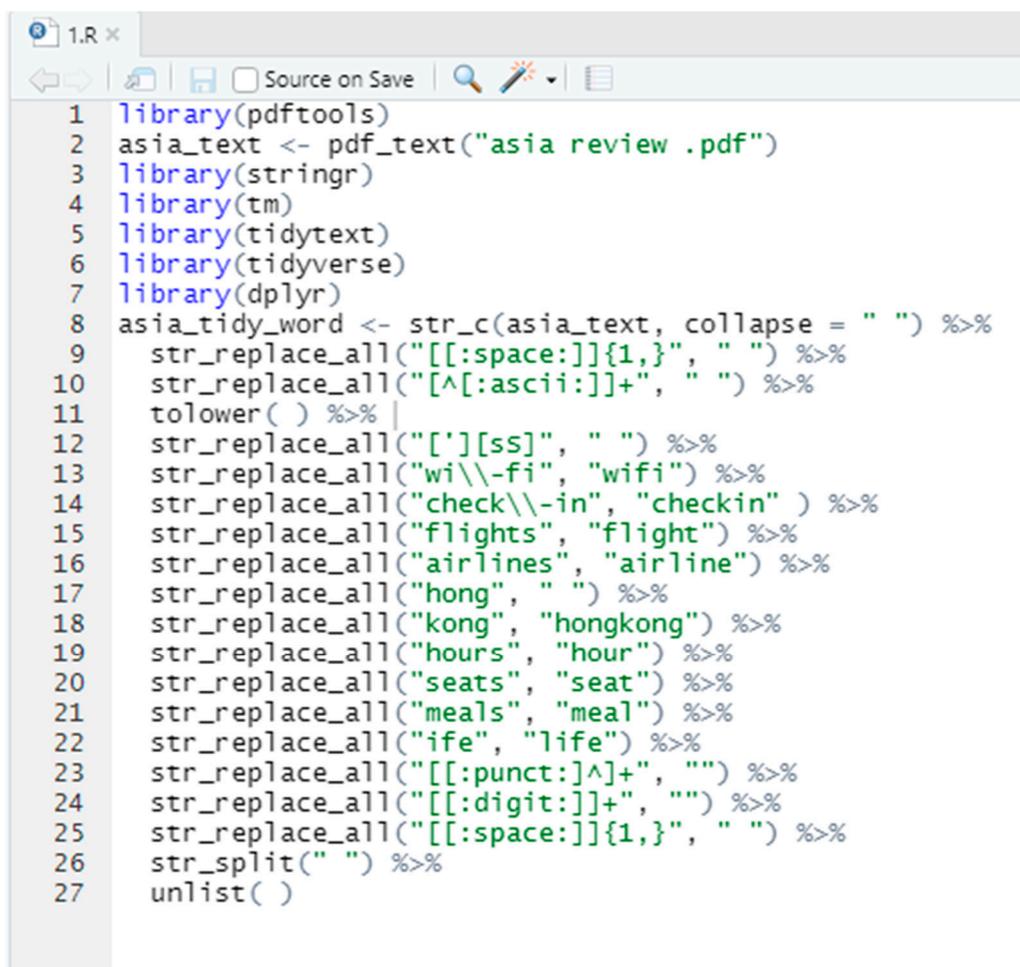


Figure 1. Research procedure.

The TM library was used as a preprocessing stage of the data. The collected online review data is organized into excel files, and only text data is converted to pdf files. A library was installed to enable the use of pdf files, assigning the object name to 'asia_text' and preprocessing it. Function packages required for preprocessing and preanalysis steps utilize stringr, tm, tidytext, tidyverse, and dplyr. First, the stripping white space was replaced with one blank that appeared more than one in a row. Additionally, because English words have upper and lower case letters and can be analyzed separately, they are unified into

lower case letters. Meaningless number expressions were removed, and sentence codes and special characters were removed. Finally, stopwords were removed. There are two types of stopwords dictionaries that can conveniently remove English words from the R program: 'en' and 'smart'. Among them, 'en' contains 174 words of words, and 'SMART' contains 571 words [47]. Therefore, the 'SMART' dictionary of words was used to remove it from this study.

Preprocessed data should go through the process of transforming data structures to proceed with topic modeling analysis during the data analysis phase. To this end, the data were transformed into documents and word matrices using the 'DTM' function. The final step is a text mining analysis step, where we performed a morphological analysis of the text using refined documents with term removal earlier. Specifically, a tokenization process was used to treat words as one word using stemming words. Morpheme is a basic analysis of the word or morpheme, the smallest unit of meaning. It uses grammatical rules or part of speech (POS), named entity recognition, spelling corrections, and word identification techniques. The combination of the morpheme analysis function commands used in this study is as Figure 2.



```

1 library(pdftools)
2 asia_text <- pdf_text("asia review .pdf")
3 library(stringr)
4 library(tm)
5 library(tidytext)
6 library(tidyverse)
7 library(dplyr)
8 asia_tidy_word <- str_c(asia_text, collapse = " ") %>%
9   str_replace_all("[[:space:]]{1,}", " ") %>%
10  str_replace_all("[^[:ascii:]]+", " ") %>%
11  tolower() %>% |
12  str_replace_all("['][ss]", " ") %>%
13  str_replace_all("wi\\-fi", "wifi") %>%
14  str_replace_all("check\\-in", "checkin") %>%
15  str_replace_all("flights", "flight") %>%
16  str_replace_all("airlines", "airline") %>%
17  str_replace_all("hong", " ") %>%
18  str_replace_all("kong", "hongkong") %>%
19  str_replace_all("hours", "hour") %>%
20  str_replace_all("seats", "seat") %>%
21  str_replace_all("meals", "meal") %>%
22  str_replace_all("ife", "life") %>%
23  str_replace_all("[[:punct:]]^+", "") %>%
24  str_replace_all("[[:digit:]]+", "") %>%
25  str_replace_all("[[:space:]]{1,}", " ") %>%
26  str_split(" ") %>%
27  unlist()

```

Figure 2. The morpheme analysis function commands.

Only the top 100 words were derived through frequency analysis. We conducted word clouds of the top 100 derived frequency-rank words. For word cloud visualization, the library functions 'wordcloud' and 'RColorBrewer' were used. We used statistical text processing techniques to estimate the probability of the emergence of topics assumed to be latent in the entire document by conducting a topic modeling analysis through structurally modified documents. In this work, we used the most widely known topic modeling latent

Dirichlet allocation (LDA) model method. Emotional analysis used a way to derive positive and negative words through emotional words left by customers in the entire document.

4. Results

4.1. Frequency Analysis

The top 100 keywords frequently mentioned in online reviews were derived as Table 2. As a result, the keywords of positive expression for airline services were derived from the top words 'good', 'comfortable', 'great', 'friendly', 'excellent', 'nice', 'clean', 'helpful', and 'pleasant'. On the contrary, the keywords for negative expression were 'return', 'delayed', 'late', 'didn't', 'poor', and 'bad'. Given that negative keywords are somewhat less than positive keywords, the online review of Asian Airlines has more positive reviews than negative ones. Additionally, many keywords indicating the regions appeared at the top, such as 'singapore', 'china', 'bangkok', 'guangzhou', 'hongkong', 'beijing', 'london', 'shanghai', 'sydney', 'manila', and 'thai'.

Table 2. The result of frequency analysis.

Rank	Word	Freq.	Rank	Word	Freq.	Rank	Word	Freq.
1	flight	20,863	35	inflight	1877	69	sydney	1013
2	seat	9798	36	back	1856	70	southern	989
3	service	7395	37	hongkong	1783	71	didn't	983
4	airline	6945	38	nice	1771	72	luggage	982
5	food	6913	39	lounge	1745	73	movies	978
6	good	6744	40	leg	1739	74	board	974
7	time	5431	41	fly	1730	75	told	967
8	crew	4640	42	served	1638	76	bit	963
9	class	4583	43	trip	1609	77	room	957
10	cabin	4547	44	checkin	1549	78	choice	952
11	staff	4529	45	delayed	1547	79	premium	948
12	hour	4275	46	beijing	1506	80	manila	942
13	meal	3841	47	clean	1502	81	onboard	927
14	business	3360	48	attendant	1490	82	quality	924
15	economy	3065	49	long	1469	83	provided	922
16	comfortable	2975	50	drinks	1428	84	small	919
17	entertainment	2943	51	london	1393	85	selection	905
18	return	2818	52	flying	1360	86	english	989
19	great	2579	53	ground	1338	87	departure	895
20	singapore	2535	54	life	1228	88	cathay	894
21	flew	2493	55	arrived	1222	89	asked	892
22	plane	2475	56	due	1198	90	price	887
23	friendly	2414	57	full	1182	91	ticket	875
24	excellent	2384	58	helpful	1175	92	pleasant	862
25	china	2341	59	minutes	1168	93	poor	861
26	airport	2339	60	airways	1140	94	booked	860
27	aircraft	2227	61	made	1112	95	thai	856
28	air	2174	62	offered	1109	96	boeing	848
29	boarding	2128	63	attentive	1097	97	times	847
30	bangkok	2081	64	efficient	1087	98	bad	842
31	check	2066	65	late	1071	99	gate	834
32	experience	2016	66	short	1035	100	left	808
33	passenger	1936	67	shanghai	1022			
34	guangzhou	1886	68	system	1014			

It can be judged that this was mentioned a lot in the review due to its high preference as an area that is operating on Asian Airlines. Among the keywords mentioned above, airline brands can also be found. The top brand was ranked 20th with 'singapore' Airlines. Additionally, along with the 'airline' keyword, 'china', 'bangkok', 'hongkong', 'airways', 'southern', 'cathay', and 'thai' were mentioned above. This can be said that customers prefer the above airline to other airlines. Finally, a number of keywords related to 'service'

The results of the topic modeling analysis are shown in Figure 4. Among the words derived from each graph of six topics, the high beta value has the most important meaning in that topic, and the researcher sets a topic name to describe the keywords contained in each topic. Topic 1 was a representative theme of 11 keywords and is named ‘In-flight meal’ which can be explained by words such as ‘flight’, ‘poor’, and ‘dessert’. Topic 2 was named ‘Entertainment’ with keywords like ‘good’, ‘great’, and ‘entertainment’ which can represent the inside of the plane. Topic 3 was named ‘Seat class’ with keywords such as ‘business’, ‘class’, ‘upgrade’, and ‘economy’ which were found to have important meanings, indicating the rating of seats on board. Topic 4 was named ‘Seat comfort’ as its subject name, including keywords like ‘seat’, ‘plastic’, ‘business’, ‘bed’, ‘space’, and ‘comfortable’. Topic 5 was named ‘Singapore Airlines’ as the subject name for ‘singapore’ keywords and ‘airlines’, ‘flight’, and ‘service’, which represent the very large beta value of the difference. The last topic 6 was named ‘Staff service’ because of the difference in importance values such as ‘cabin’, ‘crew’, and ‘food’.

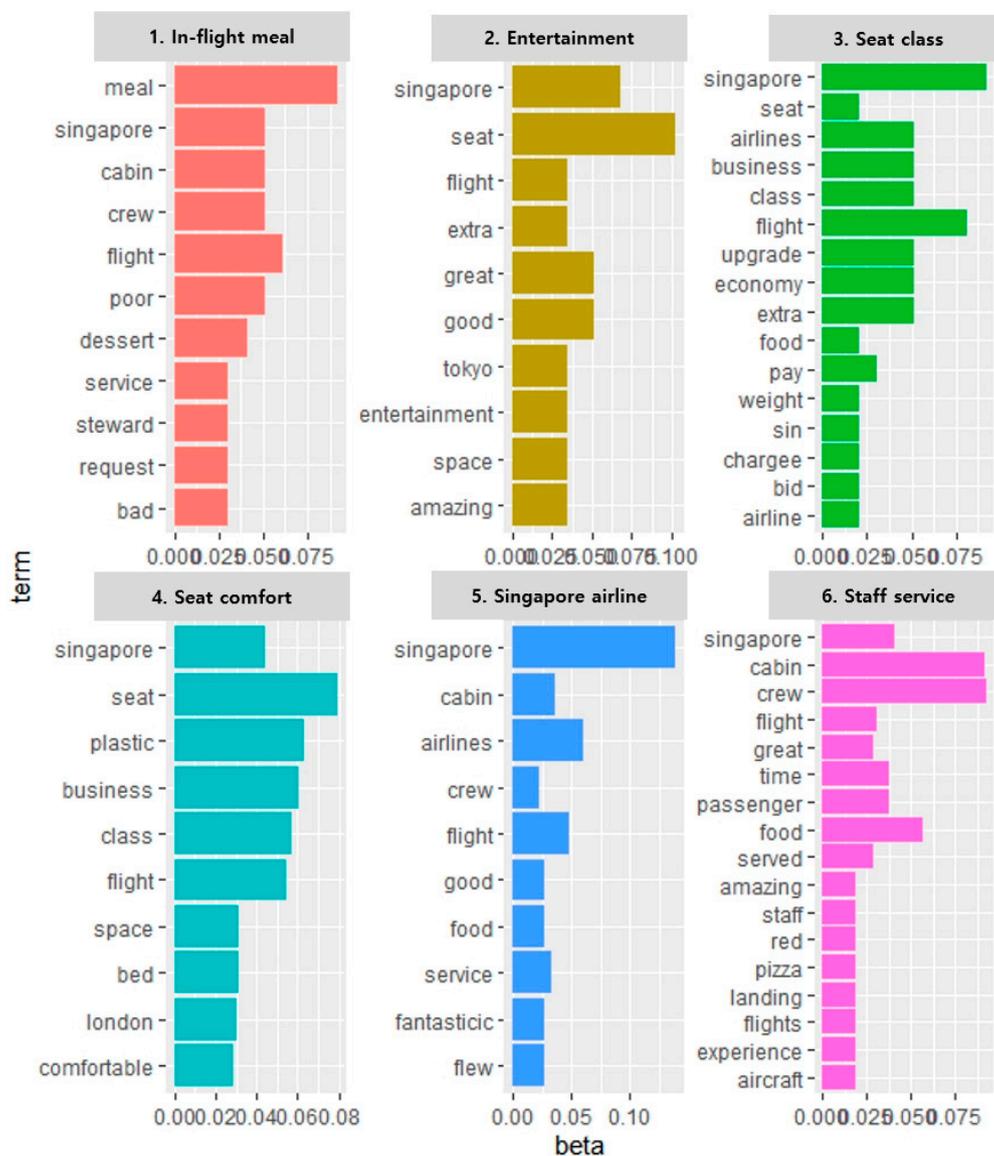


Figure 4. The result of topic modeling.

As shown above, the topics latent in full-text data of the Asian Airlines’ Online Review were grouped into six topics in total. Customers with experience using Asian Airlines can judge that the reasons for choosing to use Asian Airlines are ‘In-flight meals’,

'Entertainment', 'Seating ratings', 'Seating comfort', and 'Staff service' which are more likely to affect their purchasing intent, and especially those using Asian Airlines are very much in favor of 'Singapore Airlines'.

4.4. Sentiment Analysis

Sentiment analysis refers to a technique that classifies or quantifies emotions in text and turns them into objective information [49]. Humans use language to communicate their thoughts and feelings. If the topic modeling introduced earlier was a text mining technique that identifies the "target covered by text", the sentiment analysis is a text mining technique that estimates the "attitude contained in text". Just as the topic modeling extracts words that embody topics assumed to be inherent in the text and estimates topics, sentiment analysis also estimates the feelings inherent in the text [11].

The R program used for emotional analysis in this study is a globally accepted tool and can analyze various languages. However, it should be noted that the packages used to analyze each language are different. In the case of English, there are sentiment dictionaries and stopword dictionaries that are available to the public for sentiment analysis. This study used Opinion Lexicon, which can classify emotions as positive and negative among various sentiment dictionaries, using the 'tidytext' library for sentiment analysis of English text in R program. As a result of the analysis, 16 words were derived to indicate positive and negative.

The results showed that words expressing negative emotions, such as in Figure 5, were shown as 'poor', 'bad', 'problem', 'difficult', and 'delayed'. Therefore, it can be determined that when there is a problem, there is a negative emotion when there is a delay in service or a delay in time appointment. Words that expressed positive feelings included 'good', 'great', 'fantastic', 'excellent', and 'amazing'. These are keywords that express satisfaction after using them. 'delicious' is a word that expresses positive feelings when the airline's in-flight meal tastes good, and customers value the taste of food among what they expect from the airline. In the case of 'clean', it is the customer's desire to be clean when using the plane. As such, maintaining cleanliness can be seen as a positive emotion for customers. The word 'smiling' was derived as a word for positive emotion. This confirmed that the flight attendant's smile had a positive effect on the customer in the airline crew's service. This can be said that it is important for airlines to focus on smile education in the training of flight attendants in the future.

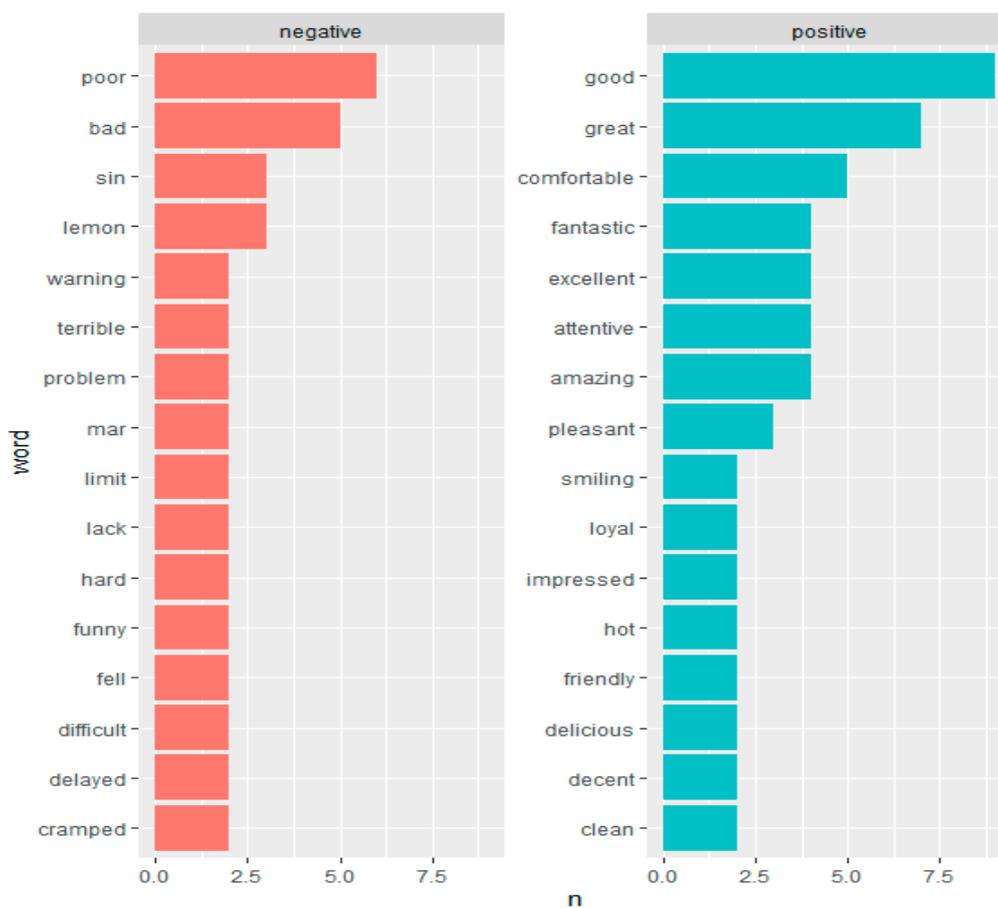


Figure 5. The result of sentiment analysis.

5. Discussion

This study used topic modeling and sentiment analysis of big data analysis to identify the needs of customers using Asian airlines as the market size of Asian airlines has become larger. By analyzing online reviews written by customers who have been experiencing Asian airlines, we explored the factors that influence the customer’s intention to purchase using Asian airlines in an exploratory approach.

Based on the results of this study, the following theoretical and practical implications were performed. First, customer needs were more clearly identified through text expressing customer opinions and feelings using online reviews from Asia Airlines to compensate for the limitations of the survey methods undertaken in many previous studies. Second, if you look at recent research trends, prior research shows various attempts to analyze big data by utilizing topical modeling and sentiment analysis. In the field of tourism and hospitality services, its utilization has also been increasing recently. In this study, customer feedback could be derived in a variety of ways by applying it to customer online reviews in the airline sector. This will be used as a marketing foundation that can be used actively in the airline service sector in the future. Third, it attempted to access the data in depth by extending it from traditional methods of analysis. This can be done by methodological understanding to establish an expanded research plan in the future.

The following are practical implications. First, frequency analysis and word cloud analysis show that among the regions where many Asian airlines operate, Singapore, Bangkok, Guangzhou, Shanghai, Hong Kong, and Beijing are frequently used and highly preferred. This could increase customer utilization if airlines use this part for marketing when promoting new flights or routes. Second, among the services provided by airlines to many customers, we could see that the comfort of the seats, the delicious in-flight meals,

and the diversity of the seat ratings were more favorable and positive for customers to choose Asian airlines. Some low-cost airlines are included among Asian airlines. However, away from the image, we could see that it was important to provide various seat upgrades and diversity in seat ratings using appropriate prices. Third, the topic modeling results confirmed that customers were very interested in Singapore Airlines among Asian airlines. In the same vein as the previous analysis, we demonstrated that there are in-flight meals, entertainment, seat ratings, seat comfort, and employee services as factors that affect customers' use of Asian airlines. It is necessary for Asian airlines to pursue diversity that will allow customers to choose from a variety of needs, using marketing to improve customer-centric services in the future.

Finally, sentiment analysis refers to negative expressions about factors that make customers feel less than expected, which negatively affects future re-use of Asian airlines. Furthermore, it will be an obstacle to developing the airline market. The most important part of it is speed. A systematic service management system must be established and operated in order to achieve the goal of services to be delivered quickly. There should also be a system that can quickly grasp the needs of customers and meet them quickly. Service training should be actively encouraged for employees to maintain consistency in service delivery. Customers expect a comfortable, clean and delicious meal to be maintained. The smile of the employee also has a positive effect on the customer, so this is also a part of the need for employee service training.

Recently, more and more studies have been undertaken using big data analysis techniques in the field of hospitality. In addition, various attempts are being made away from the existing survey methods. Representatively, text mining techniques are actively used. Although quantitative research has been mainly used, studies are actively underway to predict the future using exploratory methods while attempting to analyze using text online. Many of them use websites or social media data. In this study, it is meaningful in that the reviews left by experienced customers are derived through exploratory methods to uncover insights, predict the future, and present directions to move forward. By attempting to analyze online review data in the airline sector through topic modeling and emotional analysis, a recently actively researched text mining analysis technique, it is meaningful to present diversity in research methods in the future [33,50]. This study derives meaningful implications by analyzing what customers feel satisfied and dissatisfied with. It also seeks to provide an opportunity to assist airlines in their management activities and decision-making process, and to provide key data and analysis methods that are useful for their management.

This study has the following limitations. This study was conducted based on online reviews. There is no demographic information that many researchers point out about online reviews. Studies show that this is similar to or superior to traditional sampling, as the number of Internet and mobile users has increased in recent years, and the sample of Internet users is gradually becoming a whole population. However, since the limitations have not been fully resolved, future research may develop into a more feasible study if additional user information from the demographic site can be collected and utilized.

Author Contributions: J.-K.J. and H.-S.K. designed the research model. H.-J.K. analyzed online review data, and H.-J.K. and H.-J.B. wrote the paper. H.-J.B. was in charge of review and editing the paper. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by the Ministry of Education of the Republic of Korea and the National Research Foundation of Korea (NRF-2019S1A5A2A03049170).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Data available in a publicly accessible repository.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Heriyanto, D.S.N.; Putro, Y.M. Challenges and Opportunities of the Establishment ASEAN Open Skies Policy. *Padjadjaran J. Ilmu Huk. (J. Law)* **2019**, *6*, 466–488. [CrossRef]
2. Chen, L.; Li, Y.Q.; Liu, C.H. How airline service quality determines the quantity of repurchase intention-Mediate and moderate effects of brand quality and perceived value. *J. Air Trans. Manag.* **2019**, *75*, 185–197. [CrossRef]
3. Gupta, H. Evaluating service quality of airline industry using hybrid best worst method and VIKOR. *J. Air Transp. Manag.* **2018**, *68*, 35–47. [CrossRef]
4. Han, H.; Yu, J.; Kim, W. Environmental corporate social responsibility and the strategy to boost the airline's image and customer loyalty intentions. *J. Travel Tour. Mark.* **2019**, *36*, 371–383. [CrossRef]
5. Balakrishnan, B.K.; Dahnil, M.I.; Yi, W.J. The Impact of Social Media Marketing Medium toward Purchase Intention and Brand Loyalty among Generation Y. *Procedia-Soc. Behav. Sci.* **2014**, *148*, 177–185. [CrossRef]
6. Alnsour, M.; Ghannam, M.; Alzeidat, Y. Social media effect on purchase intention: Jordanian airline industry. *J. Internet Bank. Commer.* **2018**, *23*, 1.
7. Kim, S.; Kandampully, J.; Bilgihan, A. The influence of eWOM communications: An application of online social network framework. *Comput. Human Behav.* **2018**, *80*, 243–254. [CrossRef]
8. Lyberg, L.E.; Weisberg, H.F.; Wolf, C.; Joye, D.; Smith, T.; Fu, Y.-C. Total Survey Error: A Paradigm for Survey Methodology. In *The SAGE Handbook of Survey Methodology*; SAGE Publications Pvt Ltd.: Thousand Oaks, CA, USA, 2016; pp. 27–40.
9. Basias, N.; Pollalis, Y. Quantitative and qualitative research in business & technology: Justifying a suitable research methodology. *Rev. Integr. Busi. Econ. Res.* **2018**, *7*, 91–105.
10. Ban, H.-J.; Kim, H.-S. Understanding Customer Experience and Satisfaction through Airline Passengers' Online Review. *Sustainability* **2019**, *11*, 4066. [CrossRef]
11. Liu, B. Sentiment Analysis and Opinion Mining. *Synth. Lect. Hum. Lang. Technol.* **2012**, *5*, 1–167. [CrossRef]
12. Alaei, A.R.; Becken, S.; Stantic, B. Sentiment Analysis in Tourism: Capitalizing on Big Data. *J. Travel Res.* **2017**, *58*, 175–191. [CrossRef]
13. Ali, F.; Kwak, D.; Khan, P.; El-Sappagh, S.; Ali, A.; Ullah, S.; Kwak, K.S. Transportation sentiment analysis using word em-bedding and ontology-based topic modeling. *Knowl. Based Syst.* **2019**, *174*, 27–42. [CrossRef]
14. Park, E.; Kang, J.; Choi, D.; Han, J. Understanding customers' hotel revisiting behaviour: A sentiment analysis of online feedback reviews. *Curr. Issues Tour.* **2018**, *23*, 605–611. [CrossRef]
15. Tran, T.; Ba, H.; Huynh, V.-N. Measuring Hotel Review Sentiment: An Aspect-Based Sentiment Analysis Approach. In *Proceedings of the Computer Vision*; Springer International Publishing: Cham, Switzerland, 2019; pp. 393–405.
16. Nakayama, M.; Wan, Y. The cultural impact on social commerce: A sentiment analysis on Yelp ethnic restaurant reviews. *Infor. Manag.* **2019**, *56*, 271–279. [CrossRef]
17. Knorr, A. Big data, customer relationship and revenue management in the airline industry: What future role for frequent flyer programs? *Rev. Integr. Bus. Econ. Res.* **2019**, *8*, 38–51.
18. Thet, T.T.; Na, J.-C.; Khoo, C.S. Aspect-based sentiment analysis of movie reviews on discussion boards. *J. Inf. Sci.* **2010**, *36*, 823–848. [CrossRef]
19. Arjomandi, A.; Dakpo, K.H.; Seufert, J.H. Have Asian airlines caught up with European Airlines? A by-production efficiency analysis. *Transp. Res. Part A Policy Pr.* **2018**, *116*, 389–403. [CrossRef]
20. Lee, J.W.; Yoon, S.Y. Cross-Border Joint Venture Airlines in Asia: Corporate Governance Perspective. *Eur. Bus. Organ. Law Rev.* **2020**, *21*, 709–729. [CrossRef]
21. International Air Transport Association. Industry Statistics Fact Sheet. Available online: https://www.iata.org/pressroom/facts_figures/fact_sheets/Documents/fact-sheet-industry-facts.pdf (accessed on 20 May 2020).
22. Chung, Y.S.; Wu, C.L.; Chiang, W.E. Air passengers' shopping motivation and information seeking behaviour. *J. Air Trans. Manag.* **2013**, *27*, 25–28. [CrossRef]
23. Jeong, E.Y. Analyze of airline's online-reviews: Focusing on Skytrax. *J. Tour. Lei. Res.* **2017**, *29*, 261–276.
24. Sparks, B.A.; Browning, V. The impact of online reviews on hotel booking intentions and perception of trust. *Tour. Manag.* **2011**, *32*, 1310–1323. [CrossRef]
25. Lee, B.C.; Byun, H.J. The impact of online review on purchasing behavior: A case of hotel and resort. *J. Korean Tour. Lei.* **2014**, *26*, 59–79.
26. Aralbayeva, S.; Tao, S.; Kim, H.S. A study of comparison between restaurant industries in Seoul and Busan through big data analysis. *Culi. Sci. Hos. Res.* **2018**, *24*, 109–118.
27. Tao, S.; Kim, H.S. Cruising in Asia: What can we dig from online cruiser reviews to understand their experience and satisfaction. *Asia Pac. J. Tour. Res.* **2019**, *24*, 514–528. [CrossRef]
28. Ban, H.-J.; Choi, H.; Choi, E.K.; Lee, S.; Kim, H.-S. Investigating key attributes in experience and satisfaction of hotel customer using online review data. *Sustainability* **2019**, *11*, 6570. [CrossRef]
29. Jin, K.M.; Lee, H.R. A study on airlines brand app user's behaviour intention applied psychological decision-making process. *Inter. J. Tour. Hos. Res.* **2015**, *29*, 61–76.
30. Dellarocas, C.; Zhang, X.M.; Awad, N.F. Exploring the value of online product reviews in forecasting sales: The case of motion pictures. *J. Interact. Mark.* **2007**, *21*, 23–45. [CrossRef]

31. Sotiriadis, M.D.; VanZyl, C. Electronic word-of-mouth and online reviews in tourism services: The use of twitter by tourists. *Electron. Commer. Res.* **2013**, *13*, 103–124. [[CrossRef](#)]
32. Siering, M.; Deokar, A.V.; Janze, C. Disentangling consumer recommendations: Explaining and predicting airline recommendations based on online reviews. *Decis. Support Syst.* **2018**, *107*, 52–63. [[CrossRef](#)]
33. Gutierrez, C.E.; Alsharif, M.R.; Yamashita, K.; Khosravy, M. A tweets mining approach to detection of critical events characteristics using random forest. *Int. J. Next-Gener. Comput.* **2014**, *5*, 167–176.
34. Blei, D.M.; Ng, A.Y.; Jordan, M.I. Latent dirichlet allocation. *J. Mach. Learn. Res.* **2003**, *3*, 993–1022.
35. Kang, C.; Kim, K.K.; Choi, S. A Topic Analysis of Abstracts in Journal of Korean Data Analysis Society. *Korean Data Anal. Soc.* **2018**, *20*, 2907–2915. [[CrossRef](#)]
36. Lucini, F.R.; Tonetto, L.M.; Fogliatto, F.S.; Anzanello, M.J. Text mining approach to explore dimensions of airline customer satisfaction using online customer reviews. *J. Air Transp. Manag.* **2020**, *83*, 101760. [[CrossRef](#)]
37. Kim, S.; Park, H.; Lee, J. Word2vec-based latent semantic analysis (W2V-LSA) for topic modeling: A study on blockchain technology trend analysis. *Exp. Syst. Appl.* **2020**, *152*, 113401. [[CrossRef](#)]
38. Sutherland, I.; Sim, Y.; Lee, S.K.; Byun, J.; Kiatkawsin, K. Topic modeling of online accommodation reviews via latent dirichlet allocation. *Sustainability* **2020**, *12*, 1821. [[CrossRef](#)]
39. Lim, J.; Lee, H.C. Comparisons of service quality perceptions between full service carriers and low cost carriers in airline travel. *Curr. Issues Tour.* **2019**, *23*, 1261–1276. [[CrossRef](#)]
40. Sun, L.; Yin, Y. Discovering themes and trends in transportation research using topic modeling. *Transp. Res. Part C Emerg. Technol.* **2017**, *77*, 49–66. [[CrossRef](#)]
41. Han, G.H.; Jin, S.H. Introduction to big data and the case study of its applications. *J. Korean Data Anal. Soc.* **2014**, *16*, 1337–1351.
42. Jeong, M.S.; Shon, B.Y. Design and analysis of sentiment classification model for Korean music reviews based on convolutional neural networks. *J. Korean Data Anal. Soc.* **2018**, *20*, 1863–1871. [[CrossRef](#)]
43. Kim, J.S.; Jin, S.H. A study on the application of opinion mining based on big data. *J. Korean Data Anal. Soc.* **2013**, *15*, 101–111.
44. Wiebe, J.; Wilson, T.; Cardie, C. Annotating Expressions of Opinions and Emotions in Language. *Lang. Resour. Eval.* **2005**, *39*, 165–210. [[CrossRef](#)]
45. Esuli, A.; Sebastiani, F. SentiWordNet: A high-coverage lexical resource for opinion mining. *Evaluation* **2007**, *17*, 26.
46. Skytrax. 2020. Available online: www.airlinequality.com (accessed on 2 May 2020).
47. Feinerer, I.; Hornik, K.; Meyer, D. Text Mining Infrastructure in R. *J. Stat. Softw.* **2008**, *25*, 1–54. [[CrossRef](#)]
48. Atenstaedt, R. Word cloud analysis of the BJGP. *Br. J. Gen. Pr.* **2012**, *62*, 148. [[CrossRef](#)] [[PubMed](#)]
49. Jang, K.; Park, S.; Kim, W.-J. Automatic Construction of a Negative/positive Corpus and Emotional Classification using the Internet Emotional Sign. *J. KIISE* **2015**, *42*, 512–521. [[CrossRef](#)]
50. Reyes-Menendez, A.; Saura, J.R.; Thomas, S.B. Exploring key indicators of social identity in the #MeToo era: Using discourse analysis in UGC. *Int. J. Inf. Manag.* **2020**, *54*, 102129. [[CrossRef](#)]