

SERVICE QUALITY LEVEL EVALUATION BY ONLINE REVIEW MINING: A CASE STUDY OF INTERNATIONAL AIRPORTS

RAPHAEL GONDA, MINYOUNG KIM AND JAEHUN PARK*

Department of Business Administration
Changwon National University

20, Changwondaehak-ro, Uichang-gu, Changwon-si, Gyeongsangnam-do 51140, Korea
{20227511; 20227024}@gs.cwnu.ac.kr; *Corresponding author: pjh3479@changwon.ac.kr

Received January 2022; accepted March 2022

ABSTRACT. *The pandemic has greatly affected the airport and aviation service sectors. Airport services and facilities will be major concerns for passengers in a post-pandemic era. Online reviews should be analyzed to better understand passengers' evaluation of airport service quality. This study aims to investigate how different airport services affect customer value and satisfaction utilizing online reviews of major international airports around the world. To achieve this, the online reviews of each airport are extracted and categorized using clustering analysis. The satisfaction scores are quantified for each cluster. Through this, the key performance indicators that users consider to be important are evaluated using Term Frequency-Inverse Document Frequency (TF-IDF). In addition, benchmarking strategies are presented based on the quantified satisfaction scores of each cluster per airport. A case study evaluating the service quality performance of Incheon International Airport is conducted and benchmarking strategies are presented. The study serves as a guideline for international airports to thoroughly and more accurately understand how to improve services and facilities based on key performance indicators from online textual reviews.*

Keywords: Service quality, Text mining, Sentiment analysis, TF-IDF, KPI, Airport

1. **Introduction.** The airport and tourism industries worldwide have been hit the hardest by the ongoing COVID-19 pandemic. This is a consequence of restricted overseas travel, borders closing, and mandated quarantine protocols [1]. It has resulted in revenue loss and even bankruptcy for some companies because of the lack of passengers. International airports had invested in quarantine facilities and testing centers in combating the pandemic [2]. Hopefully, with the improving coverage of the COVID-19 vaccine around the world, some of the travel restrictions will be lifted and borders will reopen. New aviation policies are being implemented and there are signs of revival in the airport and tourism sector that could potentially increase the number of passengers adapting to the “New Normal” [3]. The temporary decline of the aviation sector could be seen as an opportunity for airports to develop strategies and plans for new and modified services to comply with post-COVID-19 requirements. Key elements of airport services include proper sanitation of facilities, sufficient and comfortable waiting areas at the gates, a wide selection of duty-free goods, and reduced waiting time for airport security and check-ins.

Customers share their experiences and feedback about airport services on online platforms. These online reviews have become essential tools in evaluating the airport's service quality and customer satisfaction. In preparation for the impending travel boom that is looming ahead, extensive research on service quality evaluation and service improvement methods is necessary. This can provide insight into modifying services that cater to a post-COVID-19 era.

This study aims to investigate how different airport services affect customer value and satisfaction utilizing online reviews of major international airports around the world. To achieve this, the online reviews of each airport are extracted and categorized using clustering analysis. The satisfaction scores are quantified for each cluster. Through this, the key performance indicators that users consider to be important are evaluated using Term Frequency-Inverse Document Frequency (TF-IDF). In addition, benchmarking strategies are presented based on the quantified satisfaction scores of each cluster per airport.

Previous studies analyzed the sentiments of airport users and the service quality provided by airports. In research from [4], aviation and non-aviation services were analyzed between major international airports to evaluate the level of customer satisfaction. In [5], Chung et al. implied the necessity to develop branding strategies that gain competitive advantage and brand value while in [6], Ryu and Park assessed the airport experiences and satisfaction level of passengers at Incheon International Airport. Compared to previous research, the main contribution of this study is to develop benchmarking strategies based on airport service quality scores evaluated using a combination of data mining and sentiment analysis.

Accordingly, the paper is organized as follows. Section 2 discusses the theoretical background of sentiment analysis and TF-IDF. Section 3 introduces the proposed framework and methodology. Section 4 presents the experimental results based on the case study of Incheon International Airport. Section 5 gives the conclusion.

2. Theoretical Background.

2.1. Sentiment analysis. Customer sentiment is a Key Performance Indicator (KPI) that measures and gauges customers' feelings toward a brand. It can tell if a customer's overall emotions were positive, negative, or neutral based on how they interacted with the brand at a specific moment. Sentiment analysis is the task of automatically extracting an individual's opinion, feeling, and emotion expressed in user-generated content [7]. Its purpose is to identify, extract, quantify, and study textual information related to customers' sentiments by analyzing subjective information and emotional states. It measures and assesses the polarity of an opinion in a review. One review may be composed of several sentences and each sentence could contain multiple opinions. Sentiment analysis obtains the polarity in opinions (i.e., positive, negative, or neutral) by comparing the number of positive words to negative words in sentences to determine their polarity. However, if the numbers are the same, the sentiment in the sentence is presumed to be neutral.

Different classifier methods and algorithms can be used to detect specific sentiments in online texts [8]. There are different kinds of sentiment analysis such as lexicon-based methods, machine learning methods, and hybrid methods. This study will utilize a lexicon-based sentiment analysis to evaluate the user satisfaction of international airport services based on the positive opinions of online reviews. The more positive opinions there are in a review, the higher the level of satisfaction. Specifically, a dictionary-based sentiment analysis, which is a method using existing dictionaries, will be applied. The advantage of using this method is the straightforward application and the requirement to manually build a new dictionary is no longer necessary.

2.2. Term Frequency-Inverse Document Frequency (TF-IDF). TF-IDF is a well-known method of weighing the importance of each word by getting the frequency of a term in a given document [9]. TF-IDF is one of the most widely used algorithms to track down key service attributes from online reviews by providing statistical information necessary in evaluating the importance of specific words. Term Frequency (TF) refers to the number of occurrences or frequency of a specific term in a document. Inverse Document Frequency (IDF) aims to allocate a smaller weightage to terms that occur frequently and increase the weightage for terms that occur infrequently in a collection of documents. It differentiates

less important documents or texts from more important ones. The frequency of words is measured by giving a bigger weightage to texts that are less common and lower weightage to more frequent texts in a corpus. This method also measures the frequency of reverse documents, which are the specific expressions to the frequency of documents.

3. **Research Methodology.** Figure 1 shows a schematic diagram of the research method. It is a modification of the level of customer satisfaction methodology suggested in the study of [10,11] to suit the purpose of this study.

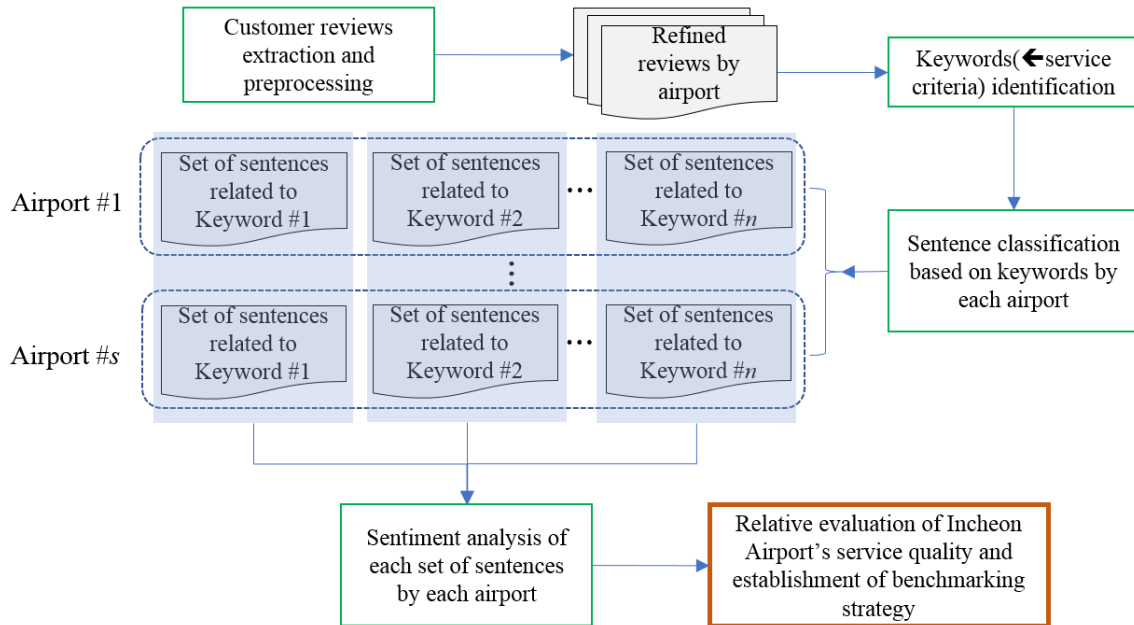


FIGURE 1. Schematic diagram of the research method

The proposed framework extracts (web-crawl) the textual reviews sourced from online reviews posted on websites. The stop words in the extracted reviews are refined to improve the accuracy of the sentiment analysis. To increase the accuracy of the analysis, unnecessary data is removed and words with similar meanings are integrated into one. Words with similar meanings (e.g., shop, shops, shopping) are integrated into the same expression (e.g., shop). For the improvement of the accuracy of service quality evaluation, sentiment analysis is performed on a sentence-by-sentence basis. Thus, the online textual reviews for each airport are subdivided into sentence units.

The KPIs that users mentioned to have great importance are identified using the TF-IDF algorithm. To identify these key service attributes, thematic words that have high relevance value (TF-IDF score) are identified. The higher the score, the more relevant the keyword (i.e., key service attribute). Sentences containing keywords are extracted and clustered. As a result, the number of clusters depends on the number of keywords extracted for each airport. The clustered sentences containing keywords will undergo sentiment analysis to measure the user's level of satisfaction.

Sentiment analysis is performed for each cluster per airport to quantify the user's service satisfaction of each KPI. The quantified service satisfaction scores are referred to as the service quality level. There are three main methods in conducting lexicon-based sentiment analysis which are the Manual approach, Dictionary-based approach, and Corpus-based approach. This study used the Dictionary-based approach, which is relatively easy and practical to implement. Positive, negative, or neutral sentences are then identified through sentiment analysis.

As a criterion for evaluating user quality satisfaction, it is assumed that positive opinions are maximized and negative opinions are minimized, the higher the satisfaction level.

In this study, based on the results of the sentiment analysis, the user service quality satisfaction value is calculated as $SQ_{ns} = \frac{NP_{ns}}{NN_{ns}}$, where NP_{ns} and NN_{ns} are the number of positive and negative sentences corresponding to the n th keyword of the s th airport, respectively.

Finally, a case study evaluating and comparing Incheon International Airport relative to the average international airport performance will be conducted. A benchmarking strategy will also be presented by recommending industry standards using the top-performing airports per service attribute.

4. Case Study.

4.1. Service quality satisfaction evaluation per airport. This study utilized the top 30 international airports in 2021 sourced from the Skytrax website. User reviews for each airport were extracted from the airport review website (www.worldairportawards.com) using R programming's "rvest" and "dplyr" packages. The number of extracted reviews varied per airport. To increase the accuracy, only 25 airports were selected for further analysis since these airports have more than 50 textual reviews. There were 7,760 reviews extracted from the 25 airports. Special symbols, numbers, incomplete characters, and non-verbal words were removed during the data preprocessing stage. Words with the same meaning but conveyed using different expressions were integrated into one word. Ultimately, the reviews were subdivided into 56,555 sentences for sentiment classification. On average, the sentiment results of each airport showed 1,035 positive, 719 negative, and 508 neutral sentences indicating that there were more positive opinions than negative ones.

Keywords that determine service satisfaction were identified through TF-IDF analysis. This analysis assumes that words frequently mentioned in reviews are factors that determine user satisfaction level. Based on each keyword, user reviews in all airports were integrated to discover which specific service attributes users considered important. Keywords with similar meanings were classified into one group, and a total of seven KPIs were classified as shown in Table 1. Focusing on each of the keywords belonging to the derived seven KPIs, clustering is performed by extracting sentences that mentioned the keywords in online reviews for each airport. Results showed that the KPIs that users care most about was Security and Waiting Time. Additionally, Wi-Fi had relatively little influence.

TABLE 1. Keyword classification based on similar meanings

KPIs	Similar meanings of the word (keyword)
Duty-Free	Shop, outlet, duty...
Waiting time	Queue, delay, wait...
Wi-Fi	Internet, data...
Sanitation	Toilet, seat...
Food	Restaurant, Café, lounge...
Traffic	Parking, bus, train, access...
Security	Staff, control, immigration...

The average positive, negative, and neutral ratio (%) sentiments of international airports were calculated as 45.72%, 31.8%, and 22.48% respectively. Figure 2 shows the results of the user service quality satisfaction level. The top five airports based on service quality satisfaction included Melbourne, Sydney, Guangzhou, Vancouver, and Hong Kong, while the bottom five are Madrid, Dusseldorf, Frankfurt, Munich, and Zurich airports.

Figure 3 shows the results of measuring the service quality level for each airport based on the seven KPIs. Texts that are in bold indicate that the service quality level is less

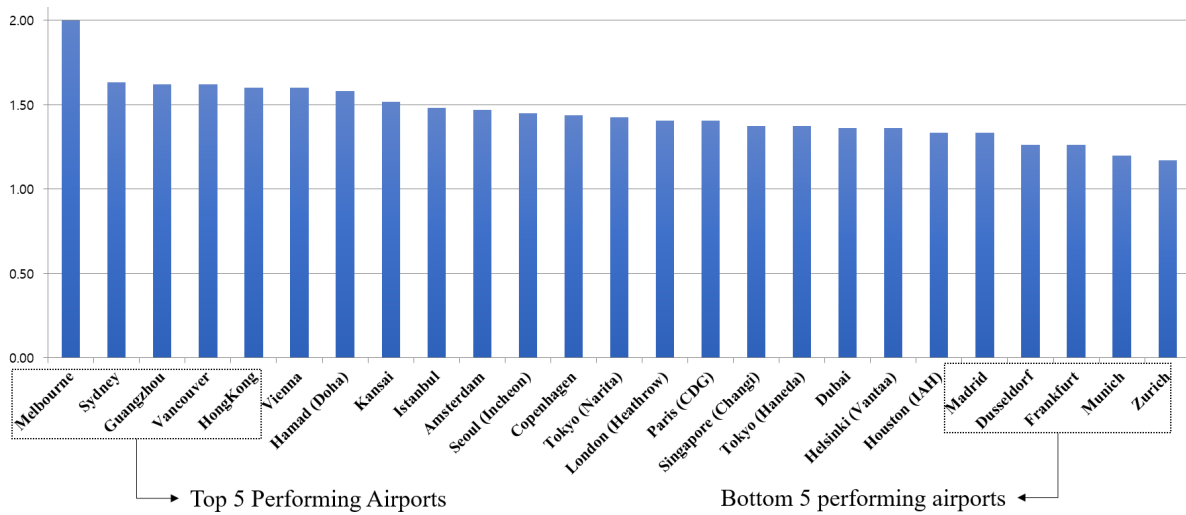


FIGURE 2. Service quality levels of the top 25 international airports in 2021

KPIs	AIRPORT												
	Amsterdam	Copenhagen	Dubai	Dusseldorf	Frankfurt	Guangzhou	Doha	Helsinki	HongKong	Houston	Istanbul	Kansai	London
Duty-Free	1.4	1.21	1.32	1.68	0.78	1.62	1.31	1.66	0.82	2.56	1.46	4.77	1.41
Waiting Time	0.81	0.94	0.77	1.16	0.71	0.85	0.89	1.87	1.43	0.48	0.78	2.67	1.17
Wi-Fi	35	2.95	5.36	0	4.56	3.77	17.25	16	0	3.5	4.89	0	7.4
Sanitation	2.28	2.91	3.16	1.25	2.12	1.65	1.74	1.79	1.39	3.05	3.05	1.16	2.28
Food	1.93	1.76	2.57	2.32	1.58	2.86	2.23	1.58	1.56	1.18	2.07	1.96	2.32
Traffic	1.07	1.42	1.87	1.44	2.4	1.94	8.75	1.75	1.71	1	2.44	1.21	1.31
Security	0.68	0.58	0.6	0.46	0.56	0.69	0.89	0.71	0.86	0.83	0.7	1	0.76

KPIs	AIRPORT												AVG
	Madrid	Melbourne	Munich	Paris	Incheon	Singapore	Sydney	Haneda	Narita	Vancouver	Vienna	Zurich	
Duty-Free	1.12	1.38	1.2	1.51	1.61	1.27	1.3	1.09	0.89	1.51	1.23	1.05	1.49
Waiting Time	0.74	0.91	0.73	0.71	2.33	2	0.57	4.75	1.24	0.78	1	0.52	1.23
Wi-Fi	8.5	6.45	7.11	3.64	11.43	8.67	3.65	0	0	0	10.13	9.57	6.79
Sanitation	1.65	2.46	1.21	2	1.41	1.85	2.95	1.81	2.12	2.95	2.5	1.16	2.08
Food	2	2.55	1.52	2.15	2.43	1.41	1.63	1.44	1.56	2.95	3.16	1.74	2.02
Traffic	0.57	1.68	1.57	1.63	0.91	1.89	1.66	1.05	1.37	2.3	1.41	1	1.81
Security	0.65	0.91	0.59	0.61	0.6	0.86	0.9	0.73	0.95	1.11	0.6	0.5	0.73

FIGURE 3. Service quality levels per KPI per airport

than 1, which means that the user’s negative sentiment is stronger than the positive one. On average, it can be seen that users generally express more positive opinions when they talk about Wi-Fi, Sanitation, and Food while Security and Waiting Time showed slightly lower satisfaction across all airports that were analyzed.

4.2. Strategies for improving service quality satisfaction. In the case of Incheon International Airport, the overall service quality level is 1.45, which is higher than that of other airports in the context of having more positive than negative sentiments. However, the service quality level is not significantly higher than that of other airports. Table 2 shows the service quality level of Incheon International Airport relative to the average level among all airports. Incheon International airport received above-average service quality scores in Wi-Fi, Food, Waiting Time, and Duty-Free. However, there was low evaluation in Sanitation, Traffic, and Security. For Wi-Fi, Food, Waiting Time, and Duty-Free, the

TABLE 2. Incheon International Airport service satisfaction comparison table

KPIs	Incheon airport's service quality level	Average service quality level	Remarks
Wi-Fi	11.43	6.79	Above average
Food	2.43	2.02	Above average
Waiting time	2.33	1.23	Above average
Duty-Free	1.61	1.49	Above average
Sanitation	1.41	2.08	Below average
Traffic	0.91	1.81	Below average
Security	0.60	0.73	Below average

service quality levels are 11.43, 2.43, 2.33, and 1.61, respectively, which are higher than the average ratings.

Concerning the Wi-Fi service attribute, Incheon International Airport is ranked second overall with an above-average service quality level of 11.43. The results suggest maintaining the current service level for this specific KPI. When looking at Waiting Time, Incheon International Airport is ranked third overall with a service quality level of 2.33, which is higher than the average. However, to provide higher service satisfaction, it is recommended to establish benchmark targets like Tokyo International (Haneda) and Kansai airports, which are the top 2 airports in the sector.

In the case of Sanitation, Incheon International Airport is ranked 20th with a below-average service quality level of 1.41. Compared to Dubai Airport, which ranked first in the category, there is a difference of 1.75. Based on the reviews, users generally showed negative sentiments on seats. To increase the rating, it is recommended to improve the sanitation of seating facilities, employ more cleaning staff, and benchmark Dubai Airport in terms of their cleaning and sanitation performance standards.

In the case of Traffic, it is ranked 24th overall with a service quality level of 0.91, which is significantly lower than the average. Compared to Doha Hamad International Airport, the top airport in the sector, there is a substantial difference of 7.84, the highest amongst the KPIs. It is presumed that the low rating is caused by the delayed dispatch times and fewer international flights in Incheon International Airport as a consequence of the COVID-19 pandemic. Other possible causes of low user evaluation could be due to the distance between public transportations and the long train queue between terminals. It is crucial to improve and reorganize the airport transportation systems to improve the service quality rating.

Lastly, about Security, Incheon International Airport is ranked 18th with a service quality level of 0.6. Compared to Vancouver Airport, the No. 1 in this category, there is a difference in service quality level of 0.5. Overall, users showed negative opinions on long security checks, immigration waiting times, and limited open counters. To improve this, it is recommended to set up fast track lanes and open more check-in counters.

5. Conclusion. Airports are planning and executing new services and seeking competitive differentiation as the world is adapting to the new normal. Online reviews are utilized as key sources of information for user requirements and service satisfaction ratings. These users generally give unsolicited and unbiased online reviews. The differences in the service quality provided by international airports affect customer value and satisfaction. Therefore, this study investigated how KPIs affect customers' satisfaction based on an online textual review analysis of services provided by major international airports.

In addition, online user reviews were compared and evaluated based on seven keywords (Duty-free, Waiting Time, Wi-Fi, Sanitation, Food, Traffic, and Security). In addition, sentiment analysis and TF-IDF analysis were conducted. Findings showed that Incheon

International Airport scored relatively low in the Sanitation, Traffic, and Security sectors. Benchmarking strategies and corrective actions are highlighted to address these problem areas to increase users' satisfaction levels. Management should assess their competitive advantage by comparing their KPIs with their competition [12].

Currently, various online service evaluation sites, including Skytrax, evaluate services based on up to five-star ratings (5-point scale). However, solely relying on 5-star ratings may neither be as comprehensive nor accurate in evaluating the level of satisfaction and the performance of KPIs. Therefore, to perform a more specific and pragmatic evaluation of airport services, online reviews written by customers could be utilized. The top 30 airports ranked by Skytrax for the year 2021 were evaluated and airports with at least 50 online reviews were used for further analysis. An analysis of 100 international airports was also conducted but the amount of data was too large to assess and would yield inaccurate results. Due to this, only the top 30 airports were considered to be the standard in evaluating the quality of airport services and facilities.

This study presents some limitations that can become opportunities for future research. The dictionary-based sentiment analysis has some restrictions. Since this study has assessed users' opinions only by word-based frequency through dictionary-based sentiment analysis, the accuracy may be reduced since stimulation of sentiments is not considered. Also, less accurate results may be yielded due to the possibility of noise from the high variation in the amount of data. To solve this problem, future research incorporating different data sources should also be conducted. Future work may consider web crawling and extract data from multiple sites in evaluating user satisfaction by matching the number of similar sentences per airport. The benchmarking strategy presented in the study did not specify any time-oriented goal. [13,14] considered a time-series model to forecast the user evaluation taking consideration of service customization and consumer innovativeness. Further research incorporating a time series of short-term to long-term goals may yield better utilization by airport management. The aviation passenger demand after the COVID-19 pandemic would likely include additional KPIs. In [15], KPIs such as quarantine and testing facilities and adhering to new aviation cultures like no-touch cultures and social distancing norms are expected to be the top concerns of airport users. Nevertheless, the results derived from this study can be a source of primary data in aiding the improvement of airport services and facilities.

REFERENCES

- [1] S. Albers and V. Rundshagen, European airlines' strategic responses to the COVID-19 pandemic (January-May, 2020), *Journal of Air Transport Management*, vol.87, DOI: 10.1016/j.jairtraman.2020.101863, 2020.
- [2] M. Kim and J. Sohn, Passenger, airline, and policy responses to the COVID-19 crisis: The case of South Korea, *Journal of Air Transport Management*, vol.98, DOI: 10.1016/j.jairtraman.2021.102144, 2022.
- [3] J. Kang, Y. Y. Jang, J. Kim, S. H. Han, K. R. Lee, M. Kim and J. S. Eom, South Korea's responses to stop the COVID-19 pandemic, *American Journal of Infection Control*, vol.48, no.9, pp.1080-1086, 2020.
- [4] S. Gitto and P. Mancuso, Improving airport services using sentiment analysis of the websites, *Tourism Management Perspectives*, vol.22, pp.132-136, 2017.
- [5] T. W. Chung, Y. J. Lee and H. M. Jang, A comparative analysis of three major transfer airports in Northeast Asia focusing on Incheon International Airport using a conjoint analysis, *The Asian Journal of Shipping and Logistics*, vol.33, no.4, pp.237-244, 2017.
- [6] Y. K. Ryu and J. W. Park, Investigating the effect of experience in an airport on pleasure, satisfaction, and airport image: A case study on Incheon International Airport, *Sustainability*, vol.11, no.17, DOI: 10.3390/su11174616, 2019.
- [7] P. Gundecha and H. Liu, Mining social media: A brief introduction, *2012 TutORials in Operations Research*, pp.1-17, 2012.

- [8] A. Iriani, Hendry, D. H. F. Manongga and R. C. Chen, Mining public opinion on radicalism in social media via sentiment analysis, *International Journal of Innovative Computing, Information and Control*, vol.16, no.5, pp.1787-1800, 2022.
- [9] G. Salton and C. Buckley, Term-weighting approaches in automatic text retrieval, *Information Processing & Management*, vol.24, no.5, pp.513-523, 1988.
- [10] J. Park, Framework for sentiment-driven evaluation of customer satisfaction with cosmetics brands, *IEEE Access*, vol.8, pp.98526-98538, 2020.
- [11] J. Park and B. K. Lee, An opinion-driven decision-support framework for benchmarking hotel service, *Omega – The International Journal of Management Science*, vol.103, DOI: 10.1016/j.omega.2021.102415, 2021.
- [12] S. Park, M. Kwak and J. Choi, Comparing competitive advantages of hotel services using opinion mining of online customer reviews: A study focusing on 5-star hotels in Seoul, Korea, *ICIC Express Letters, Part B: Applications*, vol.11, no.8, pp.743-749, 2020.
- [13] J. Lee, H. Han and S. Yoon, Air passenger demand forecasting for the Incheon airport using time series models, *Journal of Digital Convergence*, vol.18, no.12, pp.87-95, 2020.
- [14] G. Park and M. Kwak, The life cycle of online smartphone reviews: Investigating dynamic change in customer opinion using sentiment analysis, *ICIC Express Letters, Part B: Applications*, vol.11, no.5, pp.509-516, 2020.
- [15] S. H. Jin, S. J. Jeon and K. E. Kim, A study on the prediction analysis of aviation passenger demand after COVID-19, *Journal of the Korean Society for Aviation and Aeronautics*, vol.28, no.4, pp.147-153, 2020.