http://www.jsebs.org
pISSN 2288-3908
eISSN 2765-3846

The Journal of Society for e-Business Studies
Vol.26, No.4, November 2021, pp.15-36
https://doi.org/10.7838/jsebs.2021.26.4.015

머신러닝과 감성분석을 활용한 고객 리뷰 기반 항공 서비스 품질 평가

Airline Service Quality Evaluation Based on Customer Review Using Machine Learning Approach and Sentiment Analysis

전우진(Woojin Jeon)*, 이예빈(Yebin Lee)**, 금영정(Youngjung Geum)***

초 록

국제 항공 시장이 꾸준히 성장함에 따라 항공업계의 경쟁이 더욱 심화되고 있다. 경쟁 우위의 원천을 얻기 위해 서비스의 품질 평가는 필수적이며, 이에 다양한 연구에서 고객 리뷰를 바탕으로 서비스 품질을 측정하는 시도를 지속해 왔다. 그러나 고객 리뷰 데이터를 기반으로 기대와 지각 수준의 차이를 파악하고 전략적 방향을 제시하는 연구는 미흡한 실정이다. 본 연구에서는 항공사 서비스를 대상으로 차원별 중요도를 머신러닝을 통해 측정하고, 차원별 지각 수준을 감성분석을 통해 분석한다. 차원별 중요도와 지각 수준의 결과를 활용하여 항공사별 서비스의 성과를 측정하기 위한 전략 매트릭스를 제시하고, 이를 통해 각 항공사의 품질 분석을 수행한다. 본 연구는 항공사의 고객만족을 결정하는 중요한 요인을 파악하는 동시에, 각 항공사의 현재 서비스 수준을 파악하는 틀을 제시함으로써 서비스 품질 평가의 중요한 도구로 활용될 수 있다.

ABSTRACT

The airline industry faces with significant competition due to the rise of technology innovation and diversified customer needs. Therefore, continuous quality management is essential to gain competitive advantages. For this reason, there have been various studies to measure and manage service quality using customer reviews. However, previous studies have focused on measuring customer satisfaction only, neglecting systematic management between customer expectations and perception based on customer reviews. In response, this study suggests a framework to identify relevant criteria for service quality management, measure the importance, and assess the customer perception based on customer reviews. Machine learning techniques, topic models, and sentiment analysis are used for this study.

본 연구는 교육부 및 한국연구재단의 지원을 받아 수행된 연구임(2020R1I1A2070429).

* First Author, Master Course of Datascience, Seoul National University of Science and Technology (dnwls6314@ds.seoultech.ac.kr)

** Second Author, Master Course of Datascience, Seoul National University of Science and Technology (yebin@ds.seoultech.ac.kr)

*** Corresponding Author, Associate Professor, Department of Industrial & Systems Engineering, Seoul National University of Science and Technology(yjgeum@seoultech.ac.kr)

Received: 2021-07-06, Review completed: 2021-09-28, Accepted: 2021-10-05

This study can be used as an important strategic tool for evaluating service quality by identifying important factors for airline customer satisfaction while presenting a framework for identifying each airline's current service level.

키워드 : 항공 서비스, 품질 평가, 텍스트 마이닝, 토픽 모델링, 머신 러닝
Airline Service, Quality Evaluation, Text Mining, Topic Modeling, Machine Learning

1. 서 론

2000년대 등장한 저비용항공사의 성장으로 인해 항공시장이 크게 성장하고 있으며, 이에 따라 항공사 간 경쟁 역시 크게 치열해지고 있다[11]. 이러한 경쟁에서 높은 수준의 고객 만족을 달성하고 유지하는 것은 경쟁 우위를 점하기 위해 매우 중요한 활동으로 간주되어 왔다[18]. 특히 서비스의 품질 평가는 경쟁 우위의 잠재적인 원천을 얻기 위해 필수적이며, 이에 관련 연구가 다양하게 진행되고 있다[3].

서비스 품질 평가에 관련한 대부분의 기존 연구들은 고객의 서비스에 대한 기대와 지각 수준의 차이를 측정함으로써 고객 만족도를 측정한다. 이러한 접근에서는 고객이 중요하다고 생각하는 기준이 서비스로부터 충족되면 고객이 해당 서비스에 만족한다는 것을 가정한다. 대표적인 서비스 품질 평가 모형인 SERVQUAL은 서비스 품질에 영향을 미치는 5가지 차원을 유형성(Tangibles), 신뢰성(Reliability), 대응성(Responsiveness), 보증성(Assurance), 공감성(Empathy)으로 정의하고, 서비스의 품질을 서비스에 대해 고객이 기대하는 수준과 실제 제공받은 서비스에 대해 인지하는 수준 사이의 불일치 정도로써 설명하는 기법이다[22]. 이러한 SERVQUAL은 서비스의 품질 평가에 근간이 되는 기법으로 지금까지 폭넓게 활용되어 왔다[12].

오랜 시간동안 항공 서비스의 품질 평가는 설문조사 및 인터뷰를 통해 수행되어 왔으며, 이러한 접근은 가장 전통적인 접근이라 볼 수 있다[24]. 그러나 이러한 품질 평가 방식은 고객의 기억과 경험의 기간이 오래될수록 정확한 측정이 어려울 수 있고[16], 분석에 오랜 시간이 걸리며 부정확할 수 있다. 예를 들어 적은 설문 인원과 설문조사의 잘못된 항목 또는 잘못된 조사방식으로 인해 편향된 결과를 가져올 수 있다[5].

이러한 이유로 최근 온라인 고객 리뷰를 통한 품질평가가 활발하게 수행되어 왔다. 온라인 고객 리뷰는 고객이 실시간으로 받는 서비스에 대한 인식정보를 포함하므로 서비스 품질 평가에 중요하게 사용될 수 있다[17]. Lim and Lee[9]의 연구에서는 온라인 고객 리뷰로부터 토픽모델링을 이용하여 추출한 고객들의 서비스 품질에 대한 지각 특성을 SERVQUAL의 5가지 차원인 responsibility, assurance, tangibles, empathy, responsiveness과 맵핑시켜 full service carriers와 low cost carriers의 서비스 품질 인식을 비교하였고, 항공사 유형별 고객들의 감성 비교분석도 수행하였다. Park et al.[25]의 연구에서는 온라인 고객 평점 데이터를 이용하여 항공 서비스 속성들을 긍정적 또는 부정적 평점에 영향을 주는 satisfier와 dissatisfier로 분류하였고, 서비스에 대한 고객의 만족도에 영향을 줄 수 있는 속성들을 탐색하

였다. Punel et al.[26]은 온라인 고객 리뷰와 평점을 이용하여 항공 서비스 중 스카이트랙스(이하 skytrax)에서 정의한 value money rating과 overall rating에 대해 서비스 이용 고객들의 국적에 따라 서비스 기대수준이 다른지 증명하였다.

그러나 기존 연구에는 다음과 같은 한계점이 있다. 첫 번째로 평가 기준 측면에서, 대부분의 연구가 사전에 정의된 평가기준을 바탕으로 서비스를 평가하는 경우가 대부분이다. 그러나 현재 평가에 활용되고 있는 평가기준의 적절성을 검토하거나, 각 평가기준에 대한 가중치를 설정하는 등 평가기준의 적절성을 다루고 있는 연구는 매우 미흡한 실정이다. 그러나 고객들은 고객 경험을 바탕으로 항공서비스에 대한 다각적 평가를 내리는 경우가 많기 때문에 고객 리뷰를 활용해 평가기준을 재정비하는 것이 반드시 필요한 것으로 보인다. 두 번째로 평가 과정 관점에서, 대부분의 연구가 고객 리뷰를 바탕으로 고객의 만족도 및 감성을 측정하는 접근을 취해 왔으나, 고객의 기대수준과 지각 수준을 종합적으로 평가하는 것은 다소 등한시해 온 것이 사실이다. 따라서 본 연구에서는 고객 리뷰 데이터를 기반으로 항공 서비스 평가 기준의 적절성을 평가하는 동시에 평가기준의 중요도를 파악하여 고객의 기대 수준을 평가하고, 동시에 감성분석을 통해 각 평가기준에 대한 고객의 지각 수준을 파악하여 전략적 방향을 제시하고자 한다.

본 연구의 구성은 다음과 같다. 제2장에서는 본 연구와 관련된 선행 연구들을 살펴본다. 제3장에서는 본 연구의 전체적인 프로세스를 설명한다. 제4장에서는 항공 서비스 리뷰를 바탕으로 분석하여 도출된 고객의 서비스 품질에 대한 기대와 지각을 측정 및 비교를 진행한다. 마지막으로 제5장에서는 연구의 결론과 의의, 한계점 및 추후 연구 방안에 대해 서술한다.

2. 이론적 배경

2.1 서비스 품질 평가 관련 연구

서비스 품질 평가는 경쟁우위를 유지하고 소비자를 만족시키기 위한 목적으로 다수의 연구에서 시도되어 왔다. 서비스 품질평가에서 가장 범용적으로 활용되는 방법은 SERVQUAL로, Parasuraman et al.[22]에서 기업의 서비스 품질에 대한 고객의 인식을 측정하기 위해 제시하였다. 이 척도는 다섯 가지 차원으로 나뉘는데, 이 다섯 가지 차원은 유형성(Tangibles), 신뢰성(Reliability), 대응성(Responsiveness), 보증성(Assurance), 공감성(Empathy)이다. 유형성은 서비스 시설, 장비 및 직원의 외양에 관한 차원이다. 신뢰성은 약속한 서비스를 믿음직스럽고 정확하게 제공하는 능력에 관한 차원이다. 대응성은 고객을 돕고 신속한 서비스를 제공하고자 하는 의지에 관한 차원이다. 보증성은 직원의 지식과 예의범절, 그리고 믿음과 신뢰를 전달하는 능력에 관한 차원이다. 공감성은 회사가 고객에게 제공하는 인간적인 관심과 배려에 관한 차원이다.

SERVQUAL은 오랜 기간 서비스 품질 평가 연구에 광범위하게 활용되어 왔으며, 대부분 설문조사를 기반으로 서비스 품질을 파악하고자 하였다. 이러한 연구들은 항공 서비스 품질 평가를 위하여 설문조사를 통해 획득한 고객의 의견을 분석하여 서비스의 품질을 측정, 평가

해 왔다[7]. Tsaur et al.[29]은 다섯 가지 서비스 품질 기준으로 유형성, 신뢰성, 대응성, 보증성, 공감성을 이용해 항공 서비스 품질을 조사하였다. Hussain et al.[9]는 SERVQUAL 모델을 통해 아랍에미리트 기반 항공사의 서비스 품질, 브랜드 이미지 및 고객 만족 간의 연계성을 확인한 바 있다.

Park[23]는 11개의 서비스 품질 차원을 이용하여 한국 및 호주 항공사의 서비스에 대한 품질 인식 조사를 수행하였다. 해당 연구에서 이용한 11가지 서비스 품질 차원은 in-flight service, reservation-related service, airport service, reliability, employee service, flight availability, perceived price, passenger satisfaction, perceived value, airline image, overall service quality였다. 그리고 해당 서비스 품질 차원들이 항공사, 좌석 등급, 항공 서비스 이용빈도에 따라 유의미한 차이가 있는지 ANOVA, t-test를 통해 확인하였다. Pakdil and Aydın[21]은 8개의 서비스 품질 차원을 이용하여 터키 항공사의 품질 측정을 시도하였다. 해당 연구에서는 employees, tangibles, responsiveness, reliability and assurance, flight patterns, availability, image, empathy로 항공 서비스 품질 차원을 정의하였다. 그리고 가설 검증을 통해 항공사의 서비스 품질을 측정하였다.

Bari et al.[2]은 기존의 SERVQUAL을 항공 서비스 품질에 활용한 AIRQUAL을 제안하였다. AIRQUAL은 항공사 유형(Airline tangibles), 터미널 유형(Terminal tangibles), 개인(Personnel), 공감(Empathy), 이미지(Image)의 다섯 가지 측면으로 구성되어 있다. Ekiz et al.[6]과 Nadiri et al.[20]는 AIRQUAL을 사용해 항공 서비스에 대한 승객의 만족도에 관한 연구를 진행하였고, 두 연구 모두 AIRQUAL의 차원이 승객의 만족도에 크게 영향을 끼치는 것을 통계적으로 확인하였다. 하지만 두 연구 모두 설문조사를 통해 데이터를 수집하였기 때문에 데이터 수의 한계가 존재한다.

2.2 리뷰 데이터를 기반으로 연구목적별 서비스 차원을 사용한 서비스 품질 연구

최근에는 설문조사 방식 대신 온라인 고객 리뷰 데이터를 적극적으로 활용해 설문조사 방식이 갖는 한계점을 극복하고자 하는 연구가 시도되어 왔다[8]. Jeong[10]은 항공사 서비스 품질 평가기관인 skytrax에서 5개 항공사에서 총 1,565개의 리뷰를 바탕으로 항공 서비스 품질이 추천의도에 미치는 영향관계를 밝힌 바 있다. 또한 Korfiatis et al.[15]는 TripAdvisor에서 557,208개의 항공 서비스 이용 고객들의 리뷰를 수집하였다. 수집한 리뷰들을 이용해 Structural Topic Models(STM)을 통해 항공 서비스 지표 중 고객의 만족도에 영향을 크게 미치는 지표를 찾았고, 저가 항공사의 성공 요인을 밝혀내고자 하였다.

이렇듯 대다수의 연구들이 고객 리뷰를 바탕으로 다양하게 항공 서비스 품질을 측정하였으나, 대부분 연구들이 고객 만족을 평가하거나 서비스 품질이 추천의도에 미치는 관계를 파악하는 등 기대수준 또는 지각수준 중 한 가지 관점에 초점을 맞추고 있다. 뿐만 아니라 고객 리뷰를 통해 항공 서비스의 평가 기준에 대한 적절성을 검토하거나 각 평가기준에 대한 고객의 기대수준과 지각수준을 종합적으로 평가하기 위한 체계적 접근이 부족한 것으로 보인다.

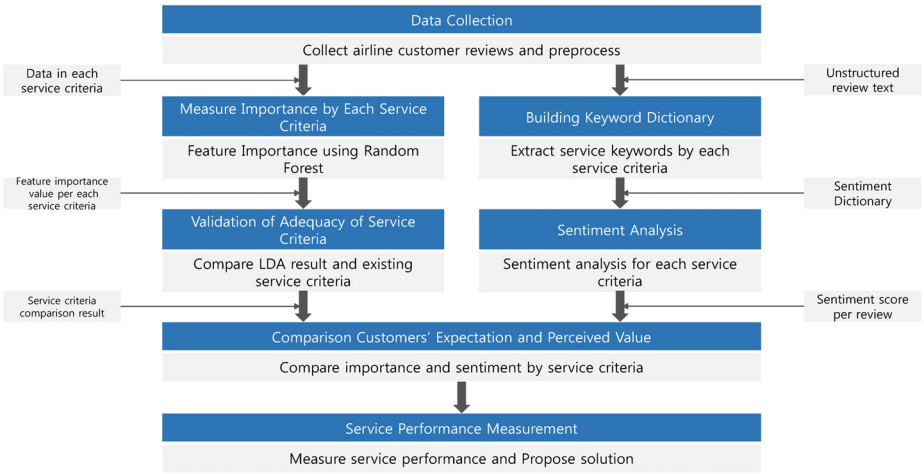
3. 연구 방법

본 연구의 절차는 <Figure 1>과 같다. 먼저 분석에 사용할 데이터를 위해 항공 서비스 이용 고객의 리뷰 데이터를 수집하고 분석하기에 앞서 전처리를 수행한다.

본 연구에서는 고객의 서비스 품질에 대한 기대와 지각 측정을 크게 두 가지 방법으로 병렬적으로 수행한다. 이를 위해 기대수준에 관련해서는 각 서비스 평가기준별 중요도, 지각수준에 대해서는 각 서비스 평가기준에 대한 만족도를 측정한다.

먼저 서비스 평가기준 관련해서, 머신러닝 기법을 바탕으로 각 서비스 평가기준을 측정한다. 본 연구에서는 이를 위해 Random forest classifier를 기반으로 한 feature importance를 측정한다. Random forest classifier는 의사 결정 나무의 해석력과 현대 학습 알고리즘의 성능을 모두 가지고 있고, 변수의 속성에 큰 영향을 받지 않는다는 장점이 있기 때문에 본 연구에서 중요도를 측정하기 위한 주요 방법론으로 활용하였다[1]. 이를 통해 항공 서비스 평가기준별 중요도를 파악해 서비스 품질에 대한 기대의 proxy로 사용할 수 있다. 또한 잠재 디리클레 할당(Latent Dirichlet Allocation, 이하 LDA)을 통해 현재 제공되고 있는 서비스 평가기준의 적절성을 검증하고자 한다. LDA를 통해 고객 리뷰로부터 몇 가지 토픽을 추출하고, 추출된 토픽이 현재 제공되고 있는 평가기준과 부합하는지를 파악하고자 한다.

서비스 지각수준 관련해서, 수집한 데이터 중 비정형 데이터인 항공 서비스 이용 고객의 리뷰를 바탕으로 각 서비스 평가기준에 적합한 서비스 키워드를 도출하고 Word2Vec으로 임베딩해 앞서 도출된 키워드와 코사인 유사도가 높은 속성어들을 얻어서 항공 서비스 평가기준별 속성 사전을 구축하고, 각 서비스 차원에 대한 감성분석을 수행한다. 이를 통해 항공 서비스를 이용한 뒤 고객이 지각한 서비스 품질에 대한 지각의 proxy로 사용할 수 있다. 마지막으로 병렬적으로 구한 항공 서비스 평가기준별



〈Figure 1〉 Overall Process of the Research

중요도와 감성점수를 비교함으로써 항공 서비스에 대한 기대와 지각의 비교분석을 수행한다. 비교한 결과를 이용해 항공 서비스의 성과를 측정할 수 있고, 개선방향을 제시할 수 있다. 다만 SERVQUAL에서 제안된 기대수준과 지각수준은 동일 차원에서 값을 측정하기 때문에 직접적인 비교가 가능하나, 본 연구에서 제안하는 기대수준과 지각수준은 고객 리뷰로부터 도출되기 때문에 직접적으로 두 지표를 비교하기는 어렵다. 따라서 본 연구에서는 각 평가기준별로 기대수준의 상대적 차이를 비교하고, 지각수준 역시 도출된 각 평가기준 내에서 상대적으로 분석하는 접근을 취한다.

3.1 데이터 수집

항공사 서비스 품질 사이트인 skytrax (https://www.airlinequality.com)에서 고객이 직접 작성한 데이터를 수집한다. Skytrax는 항공사 및 공항에 대한 평가의 신뢰도 및 인지도 면에서 전문적이고 권위있는 기관으로 평가받고 있으며 skytrax에서 직접 평가한 항공사 및 공항의 평가 결과는 산업현장의 서비스 평가 지표로 활용되기도 한다. 수집하는 데이터는 <Table 1>과 같다. 데이터는 비정형 데이터인 'content'(리뷰 텍스트)와 정형데이터인 'overall rating'(전반적인 서비스에 관한 5점 척도), 'seat comfort rating'(좌석에 관한 5점 척도), 'cabin staff rating'(승무원에 관한 5점 척도), 'food beverages rating'(기내식에 관한 5점 척도), 'inflight entertainment rating'(기내 오락에 관한 5점 척도), 'ground service rating'(지상 서비스에 관한 5점 척도), 'value money rating'(가성비에 관한 5점 척도), 'recommended'(항공사 추천 여부, binary)로 구성된다. 이 때, 5점 척도로 이루어진 데이터 속성은 높을수록 좋은 서비스를 의미하고, binary로 이루어진 데이터 속성은 1일 때 추천의사가 있는 것이고, 0이면 추천의사가 없음을 의미한다.

3.2 서비스 평가기준별 중요도 측정

앞서 수집한 데이터 중 정형 데이터에서 5점 척도를 가지는 평가항목을 항공 서비스 평가기준으로 설정하고 각각의 중요도를 측정한다. 중요도의 측정은 Random forest를 수행한 후 변수의 중요도를 permutation feature importance 에 따라 측정한다. Random forest

〈Table 1〉 Data Description

Data Form	Name	Description	Data Type
Unstructured data	content	Review text	text
Structured data	overall rating	Score for overall services	5 point likert scale
	seat comfort rating	Score for seats	5 point likert scale
	cabin staff rating	Score for crew	5 point likert scale
	food beverages rating	Score for in-flight meals	5 point likert scale
	inflight entertainment rating	Score for inflight entertainment	5 point likert scale
	ground service rating	Score for ground service	5 point likert scale
	value money rating	Score for cost-effectiveness	5 point likert scale
	recommended	Recommend the airline or not	Binary

classifier는 학습을 통해 많은 수의 개별나무들을 생성하고, 각 나무들의 분류 결과를 평균 또는 과반수투표 방식을 이용해 하나의 분류 결과로 결합한다. Random forest classifier는 bagging을 통해 개별 나무의 분산을 줄여서 정확도를 향상시킨다는 장점을 가지고 있다[1].

Input variable은 'seat comfort rating', 'cabin staff rating', 'food beverages rating', 'inflight entertainment rating', 'ground service rating', 'value money rating'으로 설정하고, target variable은 항공사 추천 여부인 'recommended'로 설정하고 학습을 수행한다. 그 후, 서비스 평가기준별 중요도를 측정하기 위해 permutation feature importance를 측정한다. Permutation feature importance는 Random forest classifier의 fitting이 끝난 뒤, input variable X_j를 하나 골라 shuffle한 뒤, shuffle 하기 전의 정확도와의 차이를 비교해 input variable X_j의 중요도를 측정하는 방법이다. 만약 input variable X_j를 shuffle한 뒤 정확도의 차이가 클수록 해당 input variable X_j는 중요도가 높다고 할 수 있고, 반대로 정확도의 차이가 작을수록 해당 input variable X_j는 중요도가 높지 않다고 할 수 있다. Permutation feature importance는 계산이 빠르며, 사용범위가 넓고, 일관된 feature의 중요도를 측정할 수 있다는 장점이 있다. 이를 도식화하면 식 (1)과 같다[1].

$$i_j = s - \frac{1}{K}\sum_{k=1}^{K} s_{k,j} \tag{1}$$

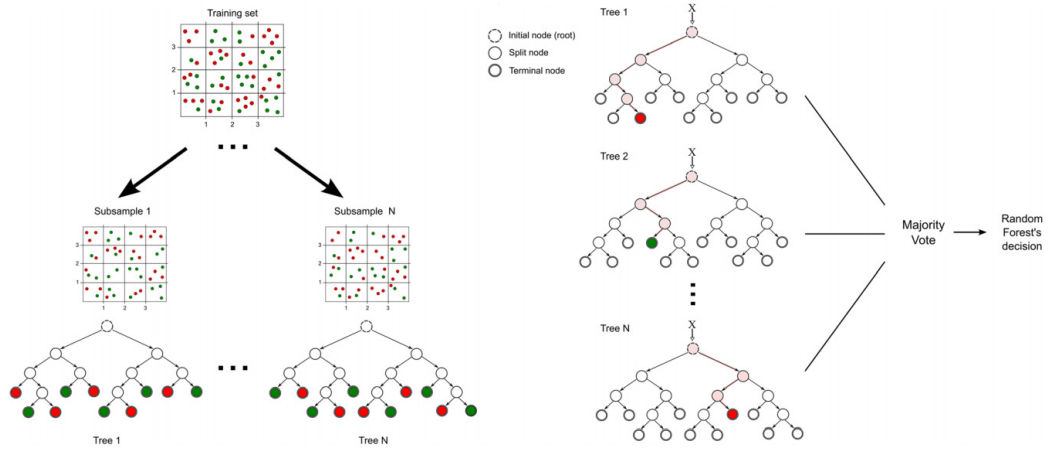
i_j = 중요도

s = reference score (기본 분류기의 정확도)

K = 반복 횟수

$s_{k,j}$ = 데이터셋 D의 column j를 무작위로 shuffle한 데이터셋 $\tilde{D}_{k,j}$의 score

본 연구는 skytrax에서 제공하는 서비스 평가기준을 사용하였기 때문에 수집한 데이터에서 고객이 직접 남긴 리뷰인 content를 활용해 skytrax에서 제공하는 서비스 유형과 리뷰에서 나타나는 토픽들을 비교하여 서비스 평가기준이 알맞게 제공되고 있는지 확인하려고 한다.

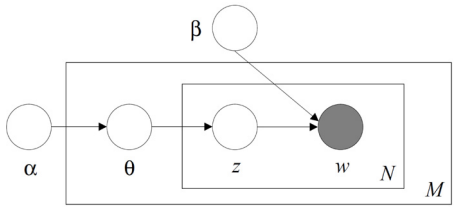


〈Figure 2〉 Random Forest〔19〕

이를 확인하는 방법으로 리뷰 데이터에 토픽모델링을 수행하여 기존에 제공되는 서비스 평가기준과 비교하고자 한다.

토픽모델링은 특정 문서의 단어들로부터 추상적인 주제를 파악하기 위해 사용되는 통계모델의 한 유형이다. LDA는 토픽모델링 방법 중 가장 대표적인 모델인데, 문서 내에 잠재되어있는 토픽을 찾아내는 알고리즘이다. 문서는 여러 개의 토픽이 포함된 것이라고 가정하고 문서 안의 단어들은 각 토픽에 속할 확률값을 가지게 된다. 그래서 설명 변수인 문서와 단어를 가지고 잠재 변수를 추론하여 문서에 속한 토픽들의 확률과 단어에 속한 토픽들의 확률을 구할 수 있다. 일반적으로 고객이 리뷰를 남기게 되면 해당 리뷰에는 고객들이 중요하게 생각하는 평가 기준 및 그에 대한 고객의 경험이 담기게 되며, 고객들은 한 가지 관점에서만 평가를 남기는 것이 아니라 다수의 평가기준을 종합적으로 평가하게 되기 때문에 하나의 리뷰에 여러 가지 토픽, 즉 평가 기준이 담긴다고 볼 수 있다. 따라서 토픽모델링은 고객의 리뷰로부터 고객이 평가하고자 하는 다수의 평가대상 및 평가기준을 담고 있는 경우가 많기 때문에 토픽 모델링은 본 연구에서 적절하게 활용될 수 있다. 특히 여러 가지 토픽 모델링 방법 중 LDA는 빠른 연산 속도를 가지고 있을 뿐 아니라 타 토픽모델링 방법론들보다 비교적 준수한 성능을 보이기 때문에 본 연구에서 사용하였다[4].

<Figure 3>에서 사용자가 설정할 수 있는 변수는 m번째 문서의 n번째 단어인 $w_{m,n}$가 유일하다. 이 변수를 가지고 α와 β를 제외한 모든 잠재 변수를 추정하게 된다. 본 연구에서 d는 리뷰 텍스트를 의미하고, z는 항공 서비스 평가 기준을 의미한다. α에 따라 Θ가 분포하게 될 Dirichlet 분포의 모양이 결정되고, η에 따라 β가 분포하게 될 Dirichlet 분포의 모양이 결정된다. 그리고 Θ에 따라 리뷰 텍스트 내에 존재하는 단어들의 토픽인 z가 결정된다. 각 단어의 토픽들을 나타내는 z와 각 단어에 대한 전체 토픽에 대한 비율 β에 따라 단어 w가 결정된다. 모델 결과는 토픽에 속하는 단어의 집합으로 도출된다.



〈Figure 3〉 LDA Model[4]

k = 토픽의 개수
M = 전체 문서의 개수
N = d번째 문서의 단어의 개수
α = 문서별 토픽 k의 Dirichlet prior weight, θ값을 결정하는 파라미터
η = 문서별 토픽 w의 Dirichlet prior weight, β값을 결정하는 파라미터
θ = 문서별 토픽의 비율
β = 토픽별 단어 w의 생성확률
z = 문서 m의 n번째 단어의 토픽(index)
w = 문서 m의 n번째 단어(문서에 관측되는 변수, index)

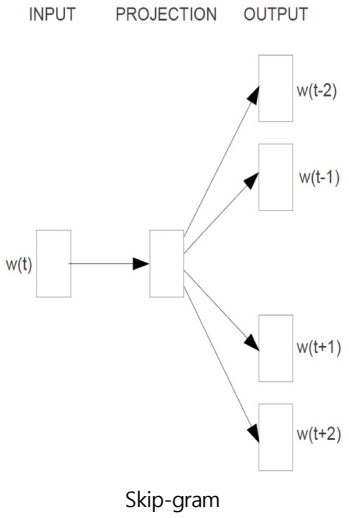
본 연구는 전체 리뷰 텍스트에서 stopwords를 제거하고 tokenize하여 LDA를 적용한다. LDA를 통해 도출된 단어 집합들을 살펴보며 단어 집합이 나타내는 주제를 명명한다. 마지

막으로 skytrax에서 제공하는 서비스 평가기준인 'seat comfort rating', 'cabin staff rating', 'food beverages rating', 'inflight entertainment rating', 'ground service rating', 'value money rating'과 비교하여 해당 평가기준들이 고객의 의견을 평가하기에 적절한지 파악한다.

3.3 서비스 평가기준에 대한 감성분석

감성분석(sentiment analysis)은 텍스트마이닝의 한 분야로 특정 문서의 긍정, 부정에 대한 감정을 추측하고 분류하는 방법이다. 감성분석은 각 문서의 최소단위인 단어의 감성극성(sentiment polarity)에 기반을 두어 이루어진다. 즉, 단어의 감성극성이 미리 정의된 감성사전을 구축한 후, 새로 주어진 문서에 출현한 단어의 감성극성에 따라 문서 전체의 감성을 분류하게 된다[14]. 하지만 감성사전을 여러 도메인에서 동시에 사용할 경우, 다양한 도메인에서의 문맥적 의미 변이, 동음이의어 사용같이 단어의 의미를 제대로 반영하지 못할 가능성이 존재한다. 예를 들어, '크다'라는 단어는 의류 도메인에서 "사이즈가 크다"와 같이 부정적인 의미로 사용될 수 있지만, 전자제품 도메인에서는 "화면이 크다"와 같이 긍정적인 의미로 사용되어 도메인별로 같은 단어가 다른 의미를 가질 수 있다. 따라서 양질의 감성사전의 구축 및 확보는 성공적인 감성분석을 위한 필수 요소라고 할 수 있고, 주제에 특화된 감성사전을 사용해 감성분석을 수행했을 때, 그 정확도가 향상됨을 확인하였다[27].

따라서 본 연구에서는 항공서비스에 특화된 감성사전을 구축하고 이를 바탕으로 항공서비스를 평가하고자 하였다. 본 연구에서는 리뷰 텍스트에서 'overall rating'의 점수가 2점 이하면 부정 리뷰라고 분류했고, 4점 이상이면 긍정 리뷰라고 분류한 뒤, 해당 리뷰들로부터 서비스 평가기준별로 특정 키워드와 코사인 유사도가 높은 단어들을 추출한다. 유사도는 Word2Vec으로 임베딩된 키워드들의 vector 간 코사인 유사도를 바탕으로 측정된다. Word2Vec은 단어를 벡터화하여 임베딩하는 방법론 중 하나이다. Word2Vec의 학습방식에는 CBOW와 Skip-gram 두 가지 방식이 있는데, 본 연구에서는 일반적으로 성능이 더 좋다고 알려진 Skip-gram 방식으로 학습을 진행한다. Skip-gram은 중심단어로 주변 단어를 예측하며 그 확률을 최대화하는 방향으로 학습하는 방식이다. 이를 시각화하면 <Figure 4>와 같다.



〈Figure 4〉 Skip-gram Model Architecture〔28〕

앞서 추출된 서비스 평가기준별 키워드와 유사도가 높은 단어들의 긍정 리뷰에서의 빈도, 부정 리뷰에서의 빈도를 통해 해당 단어의 점수를 토픽모델링의 결과로 나온 서비스 평가기준

별로 계산하고 구축된 사전의 단어들에 해당되는 점수를 맵핑하게 된다. 단어의 점수는 나이브 베이지안 확률값을 이용하고, 이를 도식화하면 식 (2)과 같다. 이때, 우도값인 $P(w_i|positive)$가 0이 되는 경우를 방지하기 위해 smoothing parameter를 추가하였다. 다음으로 리뷰 텍스트를 구축된 감성사전을 바탕으로 서비스 평가기준별 점수를 구한다. 그리고 각 리뷰가 가지는 서비스 평가기준별 점수 중 가장 큰 값을 가지는 평가기준으로 해당 리뷰를 최종적으로 분류한다.

$$P(w_i|positive) = \frac{k + count(w_i, positive)}{2k + \sum_{w \in V} count(w, positive)} \quad (2)$$

w_i = i번째 단어

$positive$ = 리뷰의 감성

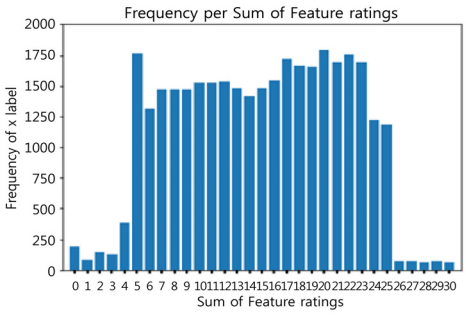
k = smoothing parameter

$count(w_i, positive)$ = 긍정 리뷰에서 단어의 빈도

4. 연구 결과

4.1 데이터 수집

본 연구는 다양한 항공 서비스에 대한 평가를 위해 항공 서비스 품질 평가기관인 skytrax에서 고객이 직접 작성한 데이터를 수집하였다. 데이터 수집 기간은 2016년 3월 1일부터 2019년 12월 31일로서 약 4년으로 설정하였다. 해당 기간동안 33,589명의 리뷰와 평점 데이터를 수집하였다. 리뷰별 'seat comfort rating', 'cabin staff rating', 'food beverages rating', 'inflight entertainment rating', 'ground service rating', 'value money rating'의 총 평점 분포를 확인해 보았을 때, 평점의 합계가 매우 낮거나 매우 높은 경우는 그 빈도가 적었고, 대부분의 경우는 분포의 중간에 위치해 있었다. 해당 분포는 <Figure 5>와 같다.



〈Figure 5〉 Distribution of Total Rating

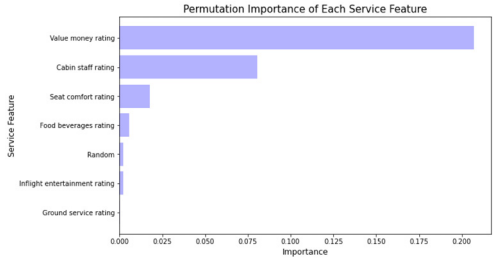
4.2 서비스 평가기준별 중요도 측정

앞서 언급된 서비스 평가기준별 중요도를 측정하기 위해 permutation feature importance, SHAP, LIME을 이용하였다. 각 서비스 평가기준의 중요도는 서로 독립적으로 계산되므로 추천 여부에 대한 개별 서비스 평가기준의 중요도를 알 수 있다. 예를 들면 Permutation feature importance는 앞서 언급한 바와 같이 Random forest classifier의 fitting이 끝난 뒤, input variable X_j를 하나 골라 shuffle한 뒤, shuffle 하기 전의 정확도와의 차이를 비교해 input variable X_j의 중요도를 측정하는 방법이다. 따라서 각 평가기준의 중요도는 직접적으로 중요도로 활용될 수 있지만, 사용되는 방법론에 따라 평가기준의 중요도 값이 상이하기 때문에 다양한 방법론에서 도출된 결과를 종합적으로 검토하여 상대적 개념의 중요도로 활용하는 것이 바람직하다.

Permutation feature importance를 측정하기 위해서는 학습이 완료된 모델이 필요하므로 중요도를 구하기에 앞서 Random forest classifier 모델의 학습을 수행하였다. 모델 학습에서 input variable은 'seat comfort rating', 'cabin staff rating', 'food beverages rating', 'inflight entertainment rating', 'ground service rating', 'value money rating'으로 정의했고, 항공사에 대한 추천여부 feature인 'recommended'를 target variable로 설정하였다. 학습이 완료된 모델을 검증 데이터로 검증을 한 결과 정확도는 약 92.69%, 정밀도는 약 93.46%, 재현율은 약 93.18%의 성능을 보였다. 모델의 학습이 끝난 후, permutation feature importance를 구한 결과는 <Table 2>와 같고, 이를 시각화하면 <Figure 6>과 같다.

〈Table 2〉 Permutation Importance of Each Service Feature

Service Feature	Importance
Value money rating	0.2070 ± 0.0057
Cabin staff rating	0.0806 ± 0.0054
Seat comfort rating	0.0176 ± 0.0008
Food beverages rating	0.0056 ± 0.0030
Random	0.0024 ± 0.0021
Inflight entertainment rating	0.0021 ± 0.0015
Ground service rating	0.0003 ± 0.0008

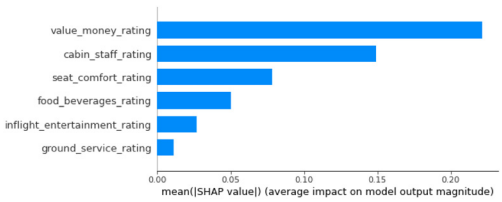


〈Figure 6〉 Permutation Importance of Each Service Feature

고객들이 항공사를 다른 사람들에게 추천하는데 영향을 미치는 서비스 평가기준별 정도는 'value money rating', 'cabin staff rating', 'seat comfort rating', 'food beverages rating', 'inflight entertainment rating', 'ground service rating'순으로 나타났다. 즉 'value money rating', 즉 가성비를 가장 중요한 차원으로 판단하였으며, 'cabin staff rating'를 그 다음으로 중요하게 판단하였다. 다음으로 'seat comfort rating'이 중요하다고 판단되었다. 'inflight entertainment rating', 'ground service rating' 등은 중요도가 난수로만 이루어진 가상 서비스 평가기준인 Random보다 중요도가 낮으므로, 고객들이 항공 서비스 평가에 크게 중요하지 않다고 생각하는 것으로 파악되었다. 또한 기존 머신러닝의 한계점이었던 개별 변수의 영향을 SHAP을 통해 측정해본 결과[13], permutation importance의 경우와 동일한 순위로 중요도가 측정되었으며 그 결과는 <Table 3>, <Figure 7>과 같다.

〈Table 3〉 SHAP Value of Each Service Feature

Service Feature	Importance
Value money rating	0.221263
Cabin staff rating	0.149348
Seat comfort rating	0.078341
Food beverages rating	0.050210
Inflight entertainment rating	0.026814
Ground service rating	0.011374



〈Figure 7〉 SHAP Value of Each Service Feature

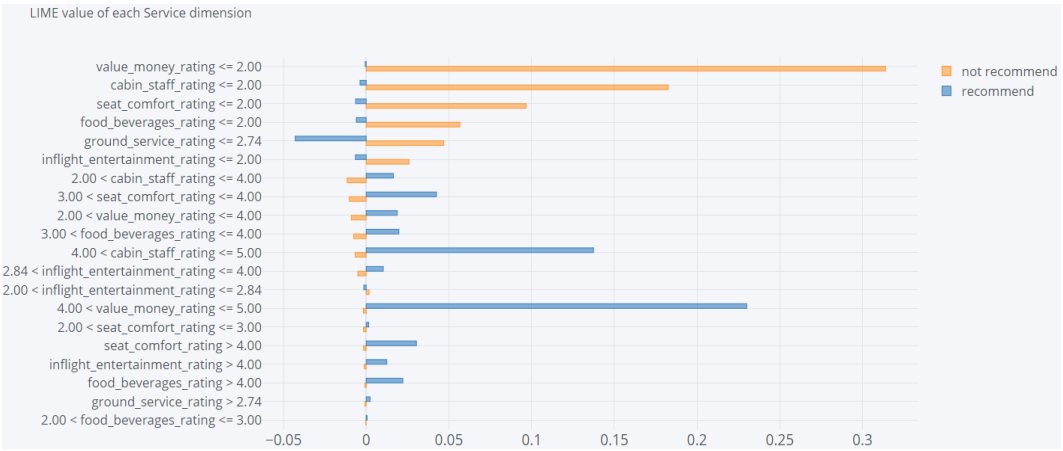
LIME으로는 각 서비스 평가기준의 평점 구간별로 추천 여부에 영향을 미치는 중요도를 측정하였다. 결과는 <Table 4>, <Figure 8>과 같다. ‘value money rating’, ‘cabin staff rating’, ‘seat comfort rating’, ‘inflight entertainment rating’ 순으로 각각 2점 이하일 때 추천하지 않는 경향이 높아짐을 알 수 있다. 이때 ‘ground service rating’은 유일하게 평점이 낮아도 추천 여부에 큰 영향을 주지 않는 것을 확인할 수 있었다. 이와 비슷하게 추천하는 경향에 영향을 미치는 순서는 ‘value money rating’, ‘cabin staff rating’, ‘seat comfort rating’, ‘inflight entertainment rating’ 순이었고, ‘ground service rating’의 평점은 높아져도 추천하는 경향이 비례하지 않았다.

다음으로 skytrax에서 제공하는 평가기준이 고객이 작성하는 리뷰에서 주로 나타나는 내용과 일치하는지 확인하기 위한 평가기준 적절성 검증을 수행하였다. 이를 위해 고객 리뷰를 바탕으로 LDA를 수행하고, 고객 리뷰가 주로 어떤 토픽으로 구성되는지 확인하였다. LDA 분석의 정확도를 높이기 위해 전처리를 수행하였으며, 공항 및 지명과 같은 불용어를 제거하였다.

LDA 분석을 효과적으로 수행하기 위해서는 LDA에 사용되는 하이퍼 파라미터에 대한 알맞은 조정이 필요하다. 하이퍼 파라미터에는 Topic(토픽의 수)과 Passes(학습 빈도)가 있다. 먼저 최적의 토픽의 수를 결정하기 위해 Passes를

〈Table 4〉 LIME Value of Each Service Feature

Not Recommend	Service Feature	Recommend
0.314025	value_money_rating ≤ 2.00	-0.0006
0.182649	cabin_staff_rating ≤ 2.00	-0.00374
0.096699	seat_comfort_rating ≤ 2.00	-0.00641
0.05666	food_beverages_rating ≤ 2.00	-0.00599
0.046956	ground_service_rating ≤ 2.74	-0.0429
0.025872	inflight_entertainment_rating ≤ 2.00	-0.00651
-0.01152	2.00 < cabin_staff_rating ≤ 4.00	0.016428
-0.01018	3.00 < seat_comfort_rating ≤ 4.00	0.042462
-0.00904	2.00 < value_money_rating ≤ 4.00	0.018801
-0.00758	3.00 < food_beverages_rating ≤ 4.00	0.019778
-0.0067	4.00 < cabin_staff_rating ≤ 5.00	0.137454
-0.00503	2.84 < inflight_entertainment_rating ≤ 4.00	0.010309
0.001744	2.00 < inflight_entertainment_rating ≤ 2.84	-0.00143
-0.00173	4.00 < value_money_rating ≤ 5.00	0.230108
-0.00165	2.00 < seat_comfort_rating ≤ 3.00	0.001503
-0.0016	seat_comfort_rating > 4.00	0.030375
-0.00121	inflight_entertainment_rating > 4.00	0.012469
-0.00083	food_beverages_rating > 4.00	0.022146
-0.00076	ground_service_rating > 2.74	0.002379
-0.00022	2.00 < food_beverages_rating ≤ 3.00	0.000561

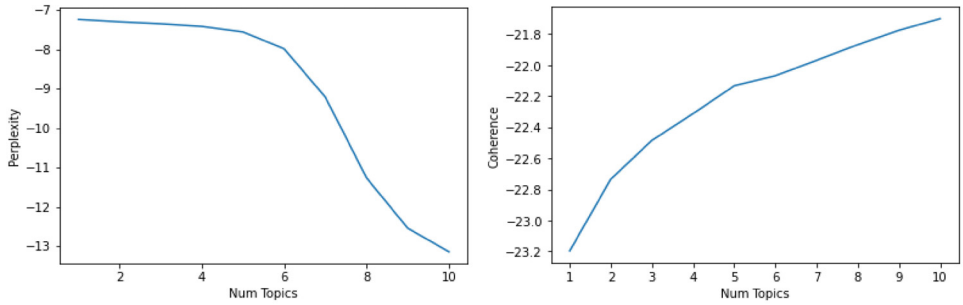


〈Figure 8〉 LIME Value for Each Service Feature Rate Interval

30으로 설정하고 Perplexity와 Coherence를 살펴보았다. Perplexity는 특정 확률 모델이 실제도 관측되는 값을 어마나 잘 예측하는지를 의미하며, Coherence는 주제의 일관성을 측정한다. Perplexity가 낮을수록, Coherence가 높을수록 적당한 Topic과 Passes를 가진다. 해당 결과는 <Figure 9>와 같다.

본 연구에서는 토픽의 수를 다양하게 설정하고 LDA 분석을 수행함으로써 다양한 결과들을 비교해보고자 하였다. 만약 토픽 내의 단어들이 하나의 주제를 나타내지 못한다고 판단되면 그 토픽은 제외하였다. 그 결과 토픽의 개수가 5개일 때, 토픽들이 명확하게 표현되었다. <Table 5>는 토픽의 개수가 5개일 때, 상위 20개의 단어를 나타낸다. LDA 특성상 토픽에 속하는 모든 단어가 해당 토픽과 직접적으로 연결되어 있지는 않지만, 상위 20개의 단어들이 나타내는 공통적인 의미가 'food beverages rating', 'service delay', 'cabin staff rating', 'seat comfort rating', 'ground service rating'를 나타내는 것을 알 수 있다.

다음으로 기존에 skytrax에서 제공하는 서비스 차원과 LDA의 분석 결과로 나온 서비스 차원과의 비교를 통해 서비스 평가기준 적절성



〈Figure 9〉 Perplexity and Coherence by Different Topic Numbers

검증을 수행하였다. 두 경우에서 공통된 서비스 차원은 'food beverages rating', 'cabin staff rating', 'seat comfort rating', 'ground service rating'였다. 하지만 skytrax에서 제공하는 서비스 차원 중 'value money rating'과 'inflight entertainment rating'은 추출된 토픽에서 찾아볼 수 없었다. 반면에 리뷰 텍스트로부터 추출된 토픽 중 'service delay'는 skytrax에서 제공하는 서비스 차원에서 제공하지 않는 새로운 서비스 차원으로 항공 서비스를 평가할 수 있는 새로운 서비스 평가기준으로 사용할 수 있는 것으로 판단되었다.

따라서 본 연구에서는 skytrax에서 제공하는 서비스 평가기준과 LDA 분석을 이용해 리뷰 텍스트로부터 추출한 토픽을 혼합해 5가지 서비스 평가기준으로 사용하였다. 5가지 평가기준은 'seat comfort rating', 'cabin staff rating', 'food beverages rating', 'ground service rating', 'service delay'이다. 기존에 skytrax에서 제공하는 서비스 평가기준은 'seat comfort rating', 'cabin staff rating', 'food beverages rating', 'ground service rating'이고, 리뷰 텍스트로부터 추출되어 새롭게 추가된 서비스 평가기준은 'service delay'이다. 특히'value money rating'기준의 경우, 앞서 언급한 대로 평가기준 중요도는 가장 중요하게 도출되었으나, 토픽 모델링을 통해 도출한 토픽에는 직접적으로 드러나지 않았다. 이는 해당 평가기준이 높은 중

〈Table 5〉 5 Topics with Top 20 Words of Reviews

	Food Beverages	Service Delay	Cabin Staff	Seat Comfort	Ground Service
	words	words	words	words	words
1	serve	wait	flight	class	flight
2	meal	gate	good	business	delay
3	drink	delay	time	seat	hours
4	offer	cancel	crew	service	airport
5	water	hotel	food	economy	time
6	breakfast	counter	friendly	food	board
7	sandwich	weather	service	lounge	check
8	coffee	refund	cabin	good	tell
9	dinner	miss	staff	first	wait
10	snack	return	return	excellent	take
11	wine	could	comfortable	cabin	would
12	meals	even	great	comfortable	hour
13	crew	crew	entertainment	sleep	arrive
14	chicken	travel	excellent	flat	gate
15	food	entertainment	clean	offer	leave
16	cold	extra	aircraft	better	minutes
17	glass	cabin	nice	great	late
18	land	make	seat	premium	staff
19	bottle	airlines	overall	quality	passengers
20	lunch	experience	inflight	selection	luggage

〈Table 6〉 Sentiment Score of Words in Each Service Criteria

Seat Comfort			Cabin Staff			Food Beverages			Ground Service			Service Delay		
words	positive	negative	words	positive	negative	words	positive	negative	words	positive	negative	words	positive	negative
seat	0.336999	0.429711	staff	0.314238	0.410307	food	0.336121	0.629607	lounge	0.051802	0.157214	wait	0.021691	0.015594
cabin	0.198514	0.399564	crew	0.223304	0.420856	wine	0.027834	0.057071	terminal	0.035348	0.055378	gate	0.121949	0.065406
legroom	0.03469	0.086372	cabin	0.198514	0.399564	snack	0.01692	0.054467	board	0.083338	0.100957	delay	0.089206	0.051081
row	0.048127	0.064299	inflight	0.072753	0.127718	breakfast	0.037213	0.074847	check	0.090578	0.130518	cancel	0.000795	3.26E-05
leg	0.116574	0.192961	english	0.024434	0.038189	water	0.073959	0.056681	tell	0.000521	0.000163	hotel	0.060522	0.023408
choice	0.050047	0.092818	captain	0.018236	0.025361	sandwich	0.013136	0.030245	gate	0.121949	0.065406	counter	0.041655	0.020087
space	0.037981	0.063322				meal	0.113393	0.168544	minutes	0.000137	9.77E-05	weather	0.024653	0.01605
room	0.077031	0.106036				drink	0.040394	0.058439	late	0.110925	0.097441	refund	0.011874	0.00153
seating	0.047579	0.071787				coffee	0.026189	0.031677	staff	0.314238	0.410307	miss	0.002276	0.001791
screen	0.032661	0.045872				dinner	0.027011	0.048476	passengers	0.001837	0.001139	waiting	0.07374	0.032394
lounge	0.051802	0.157214				meals	0.000905	0.003874	luggage	0.110377	0.088911	outbound	0.027779	0.051732
sleep	0.009571	0.01885				chicken	0.015549	0.02764	boarding	0.162316	0.19205	board	0.083338	0.100957
flat	0.017578	0.066578				glass	0.021088	0.014227				leave	0.000631	0.000163
sector	0.008583	0.022757				bottle	0.013245	0.013836				departure	0.073301	0.07081
rest	0.02405	0.025752				lunch	0.012148	0.027185				hours	0.000795	0.000163
interior	0.007212	0.020803				vegetarian	0.013849	0.01579				minutes	0.000137	9.77E-05
bed	0.006938	0.026924				juice	0.010009	0.018459				late	0.110925	0.097441
						tea	0.015878	0.019241						
						quality	0.044507	0.077582						
						bar	0.008199	0.02139						
						catering	0.008803	0.026143						
						alcohol	0.006499	0.014292						

요도에도 불구하고 고객들이 직접적으로 리뷰를 통해서 의견을 피력하는 평가기준이 아님을 의미한다.

따라서 본 연구에서는 'value money rating' 평가기준의 중요도가 가장 높았음에도 불구하고 최종 서비스 평가기준에서 제외한다. 이러한 배경에는 개념적으로 가격 대비 성능 즉 가성비는 모든 종류의 의사결정에 직접적으로 관련이 있어 많은 고객들이 항공사 서비스를 이용하면서 가격에 기반해 대안을 필터링한다는 점, 평가기준 자체가 가지는 의미가 매우 포괄적인 점, 항공사 추천 여부와 기본적으로 상관관계가 매우 높은 기준이라는 점이 고려되었다. 특히 본 연구의 경우 토픽 모델링에서 드러나는 평가기준을 바탕으로 관련 단어를 추출하여 각 토픽에 대한 감성점수를 계산하기 때문에 토픽에서 드러나지 않은 평가기준은 추후 분석에서 고려하기 어렵다는 면이 있다.

4.3 서비스 평가기준에 대한 감성분석

다음은 리뷰 텍스트로부터 추출한 각 서비스 평가기준(토픽)에 해당하는 고객의 인지 수준, 즉 감성 점수를 판단하기 위해 감성분석을 수행하였다. 먼저 각 서비스 평가기준별로 대표 키워드를 선정하고, Word2Vec을 통해 서비스 평가기준별 대표 단어와 코사인 유사도가 높은 단어들을 추출하여 키워드 사전을 구축하였다. 구축된 단어 사전을 바탕으로 최종 선정된 단어들이 긍정과 부정리뷰에서 쓰일 때의 점수를 앞서 설명한 나이브 베이지안 확률값으로 맵핑하였다. 그 결과는 <Table 6>과 같다.

이를 바탕으로 리뷰들을 미리 분리한 서비스 평가기준에 따라 분류하였다. 분류는 각 리뷰가 보이는 서비스 평가기준별 점수 중 가장 큰 값을 나타내는 평가기준으로 해당 리뷰를 분류하였다. 예를 들어 리뷰의 서비스 평가기준별 점수 중 'cabin staff rating'의 점수가 가장 높게 나오면 해당 리뷰에 대한 내용은 승무원에 관한 리뷰로 분류할 수 있다. <Table 7>은 실제 데이터 중 하나의 리뷰를 어떻게 분류하는지 보여준다. 서비스 평가기준 중 좌석에 대한 점수가 가장 높게 나왔으므로 좌석에 관한 리뷰로 분류하게 된다.

기존의 리뷰에서 'overall rating'을 기준으로 긍정과 부정으로 분류된 리뷰에서 언급되는 서비스 평가기준별 비중을 찾기 위해 전체 리뷰에 대하여 분류를 수행하였다. <Table 8>은 전체 리뷰를 분류했을 때의 결과를 나타낸다. 결과를 살펴보면 긍정리뷰에서 주로 나타나는 서

〈Table 7〉 Example of Sentiment Score by Service Criteria in Review

Excellent comfort and food, which is served on demand during the flight.
The seat converts into a flat bed at the press of a button. The seat and the pillow was ultra-comfortable. Passengers are given a set of pyjamas for red-eye flight and amenities from Giorgio Armani in a pouch. Service, however, can be improved as it is not consistent. I had to wave to get the attention of the flight attendant as no one had attended to me for more than 5 min after the attendant light was turned on. I also had to remind the flight steward to collect my glass prior to take-off. No one reminded me to stow away my tray table prior to landing.

Food Beverages	Cabin Staff	Seat Comfort	Service Delay	Ground Service
0.0211	0	0.6985	0	0

〈Table 8〉 Proportion of Service Criteria in Review

Service Criteria	Frequency in positive review	Frequency in negative review	Proportion in positive review	Proportion in negative review	Proportion in entire review
Seat Comfort	4425	3561	55.41%	44.59%	23.78%
Food Beverages	5438	4026	57.46%	42.54%	28.18%
Cabin Staff	3511	2024	63.43%	36.57%	16.48%
Ground Service	4199	4247	49.72%	50.28%	25.15%
Service Delay	659	1499	30.54%	69.46%	6.42%

비스 차원은 'Seat Comfort', 'Food Beverages', 'Cabin Staff'이었다. 반면에 부정 리뷰에서 나타나는 서비스 차원은 'Service Delay'였고, 앞서 서비스 평가기준별 중요도를 측정했을 때와 비슷하게 'Ground Service'는 긍부정 비율에 큰 차이가 없어서 중립적인 감성으로 나타났다.

4.4 서비스 평가기준에 대한 고객의 인식과 지각 비교

항공 서비스에 대한 고객들의 인식과 지각을 서비스 평가기준별로 비교해 보았다. 먼저 'Seat Comfort'의 중요도는 0.0176 ± 0.0008이었다. 긍정리뷰 비율은 55.41%였고, 부정리뷰 비율은 44.59%였다. 또한 전체 리뷰에서 해당 서비스 평가기준이 차지하는 비중은 23.78%로, 세 번째로 높은 비중을 차지한다. 즉 항공사들은 해당 서비스에 대한 평가 결과에 대해 안주하지 않고 지속적인 개선과 발전을 위해 노력해야 함을 알 수 있다. 'Food Beverages' 의 중요도는 0.0056 ± 0.0030이었다. 긍정리뷰 비율은 57.46%였고, 부정리뷰 비율은 42.54%였다. 전체 리뷰에서 해당 서비스 평가기준의 비중은 28.18%로, 전체 리뷰에서의 비중이 가장 높았다. 이는 'Seat Comfort'와 비슷한 양상으로, 그만큼 고객들이 해당 서비스에 대해 기억하고 비교적 해당 서비스 평가기준에 대해 많은 리뷰를 남기는 것을 알 수 있다.

'Cabin Staff'의 중요도는 0.0806 ± 0.0054이었다. 긍정리뷰의 비율은 63.43%였고, 부정리뷰의 비율은 36.57%, 전체 리뷰에서의 비중은 16.48%였다. 이 평가기준은 중요도 평가에서는 'Value money rating'에 이어 두 번째로 높게 나타났으나, 전체 리뷰에서의 비중이 낮게 나타났다. 이는 고객이 기내 승무원 서비스에 대해서 일반적으로 매우 중요하게 평가하지만, 실질적으로 서비스를 경험한 후 승무원 서비스에 대해서는 리뷰를 남기지 않는다는 것을 의미한다. 이는 고객이 승무원 서비스에 대해 관념적으로는 중요하게 여기지만 실질적으로 서비스를 이용한 후에는 해당 평가기준을 심각하게 고려하지 않는다는 것을 의미한다. 한 가지 눈여겨 볼 사실은 기내 승무원 서비스의 경우 타 평가기준과 달리 긍정리뷰의 비중이 크게 높다는 점이다. 즉 일반적으로 현재의 항공 서비스에서는 승무원 서비스가 충분한 수준으로 제공되고 있다는 것을 시사한다.

'Ground Service'의 중요도는 0.0003 ± 0.0008이었다. 긍정리뷰 비율은 49.72%였고, 부정리뷰 비율은 50.28%였으며, 전체 리뷰에서 차지하는 비중은 25.15%였다. 주목할 점은 이 평가기준의 경우 머신러닝을 바탕으로 측정한 중요

도는 가장 낮지만 전체 리뷰에서의 비중은 두 번째로 높았다는 점이다. 이는 고객들이 관념적으로 지상서비스를 중요하게 평가하지는 않지만, 서비스 경험 후에는 의외로 많은 고객들이 지상서비스에 대해 높은 비중으로 리뷰를 남긴다는 것을 의미한다. 따라서 지상서비스의 경우 평가기준의 중요도가 낮게 나오더라도 품질관리를 등한시하지 않고 지속적으로 관리하는 것이 필요하다. 많은 항공사가 기내 서비스에 초점을 맞춰 서비스를 제공하고 경쟁 우위를 확보하고 있는 실정이지만, 지상 서비스 역시 고객들의 서비스 만족도를 높일 수 있는 중요한 방안으로 간주될 수 있다.

마지막으로 'Service Delay'는 토픽모델링의 결과로 인해 새롭게 추가됐으므로 중요도는 계산되지 않았다. 긍정리뷰에서의 비율은 30.54%였고, 부정리뷰에서의 비율은 69.46%, 전체 리뷰에서의 비중은 6.42%였다. 즉 모든 서비스 평가기준을 통틀어 고객들이 가장 관심을 두지 않는 평가기준임을 알 수 있다. 그러나 고객들이 많이 언급하지는 않더라도, 서비스 지연의 경우 항공사 품질의 핵심적인 요인이기 때문에 지속적인 관리 및 개선 노력이 필요하다고 보인다.

5. 결 론

본 연구는 항공 서비스에서 고객의 기대와 지각 수준의 차이를 파악하고 산업내 경쟁에 도움을 줄 수 있는 전략적 방향을 제시하기 위해 고객 리뷰 데이터를 활용하였다. 머신러닝 방법론 중 하나인 Random forest를 이용하여 항공서비스 평가기준별 중요도로 고객의 기대를 측정하고 토픽모델링을 수행하여 기존에 제공되던 항공 서비스 평가기준을 검사해봤고, 기존의 서비스 평가기준 4개와 토픽모델링을 통해 새롭게 추가된 평가기준 1개를 통합해 총 5가지의 서비스 평가기준을 이용하였다. 5가지의 서비스 평가기준별로 리뷰의 감성을 분류하는 감성분석을 진행하였다. 감성분석은 항공 도메인의 특성을 반영하기 위해 항공 도메인에 맞게 새롭게 구축하여 진행하였다.

본 연구에서는 항공사의 고객만족을 결정하는 중요한 요인을 파악하는 동시에, 각 항공사의 현재 서비스 수준을 파악하는 틀을 제시함으로써 서비스 품질 평가의 중요한 도구로 활용될 수 있고, 이를 통해 고객의 기대와 지각 수준의 차이를 파악하여 전략적 방안을 제안하는 데 도움을 줄 수 있다. 특히 머신러닝을 바탕으로 도출된 각 평가기준의 중요도, 토픽 모델링을 통해 고객이 직접적으로 표현한 각 평가기준의 비중 및 특성, 각 평가기준에 대한 긍정, 부정 비율 등을 종합적으로 살펴봄으로써 항공사의 품질관리를 위해 각 평가기준을 어떤 관점으로 접근해야 하는지에 대한 전략적 시사점을 제안하였다.

본 연구에서 제안된 프레임워크에 따라 항공서비스에 대한 고객의 기대와 지각 차이를 살펴본 결과 좌석과 관련된 서비스와 기내식 서비스의 경우 지속적인 개선이 요구되었다. 특히 기내 승무원과 관련된 서비스에서 고객들의 기대도 높고 그에 알맞은 서비스가 제공되고 있음을 확인하였다. 하지만 기내 승무원 관련해서는 고객들이 서비스 경험 후 이에 관련해 리뷰를 적극적으로 남기지는 않는 것으로 파악되었으며, 이에 기내 승무원과 관련된 서비스 수준은 유지하면서 개선이 필요한 다른 서비스에 더 많은 노력을 쏟을 필요가 있다고 생각된

다. 또한 지상 서비스는 평가기준의 중요도는 낮게 도출되었으나 서비스 경험 후 많은 고객들이 이에 관해 평가를 남기는 것으로 드러났다. 따라서 낮은 중요도에도 불구하고 항공사들이 차별화된 지상 서비스를 고민하여 제공함으로써 경쟁 우위를 가질 수 있다고 판단된다. 마지막으로 서비스 지연에 대한 부분은 고객들의 관심이 가장 크지 않은 부분이었다. 그럼에도 불구하고 고객들의 서비스 불만족이 가장 큰 평가기준이기도 했다. 따라서 지상 서비스와 마찬가지로 항공사가 서비스 개선을 통해 경쟁 우위를 가져올 수 있는 서비스 평가기준이라고 보여진다.

본 연구의 한계점은 다음과 같다. 먼저 SERVQUAL에서는 기대수준과 인지수준을 동일한 차원에서 측정하기 때문에 gap 분석이 용이한 반면, 본 연구에서는 두 수준이 같은 차원에서 측정되지 않았기 때문에 기대수준 및 인지수준은 각 평가기준 내에서 상대적으로 평가되었다. 따라서 기대수준과 인지수준의 직접적인 비교가 어렵다는 한계가 있다. 또한 영어로 기록된 리뷰 데이터만 사용했기에 텍스트 분석의 특성상 모든 언어들을 분석할 수 없었고, 비영어권 고객들이 대부분 이용하는 항공사 및 언어권 차이에 따른 서비스 만족도 및 항공사의 다른 전략적 방향이 존재할 수 있다는 점이다. 또한, 텍스트 데이터에 대하여 클러스터 개수를 정하거나 이름 지을 때, 주관적인 해석을 배제할 수 없었다는 점이다. 이는 토픽모델링 분석에서 자주 겪는 문제로서 최신 알고리즘을 적용하여 주관적 해석이 들어가는 단점을 최소화해야 할 것이다. 또한 토픽모델링을 통해 추가된 평가기준에 대해서는 중요도를 평가할 수 없었다는 점이 한계점으로 생각된다. 마지막으로 본 연구는 전체 항공 서비스에 대한 분석을 실시하였으며, 개별 항공사에 따른 서비스 만족도를 분석하지 않았다. 추후 연구에서 서비스 이용 목적이나 항공사에 따른 서비스 만족도 차이를 통해 더욱 자세한 전략적 방향을 제시한다면 보다 실무적인 시사점을 제공할 수 있을 것으로 보인다.

References

[1] Andre Altmann, Laura Toloşi, Oliver Sander, Thomas Lengauer, “Permutation importance: A corrected feature importance measure,” Bioinformatics, Vol. 26, No. 10, pp. 1340-1347, 2010.

[2] Bari, S., Bavik, A., Ekiz, H. E., Hussain, K., and Toner, S., “AIRQUAL: A multiple-item scale for measuring service quality, customer satisfaction, and repurchase intention,” HOS-414 Graduation Project (Thesis), pp. 1-104, 2001.

[3] Bharadwaj, S.G., Varadarajan, P. R., and Fahy, J., “Sustainable competitive advantage in service industries: A conceptual model and research propositions,” Journal of Marketing, Vol. 57, pp. 83-99, 1993.

[4] Blei, D. M., Ng, A. Y., and Jordan, M. I., “Latent dirichlet allocation,” Journal of Machine Learning Research, Vol. 3, pp. 993-1022, 2003.

[5] Chow, C. K. W., “On-time performance,

passenger expectations and satisfaction in the Chinese airline industry," Journal of Air Transport Management, Vol. 47, pp. 39-47, 2015.

[6] Ekiz, H. E., Hussain, K., and Bavik, A., "Perceptions of service quality in North Cyprus national airline," 18th Biennial International Conference on Tourism and Hospitality Industry, pp. 778-790, 2006.

[7] Gilbert, D. and Wong, R. K., "Passenger expectations and airline services: a Hong Kong based study," Tourism Management, Vol. 24, No. 5, pp. 519-532, 2003.

[8] Han, M. and Choi, B., "Data Mining-Based Airline Service Quality Analysis," The 2019 Spring Conference of The Korea Society of Management information Systems, pp. 692-705, 2019.

[9] Hussain, R., Al Nasser, A., and Hussain, Y. K., "Service quality and customer satisfaction of a UAE-based airline: An empirical investigation," Journal of Air Transport Management, Vol. 42, pp. 167-175, 2015.

[10] Jeong, E., "Analyze of Airline's Online-reviews: Focusing on Skytrax," Journal of Tourism and Leisure Research, Vol. 29, No. 1, pp. 261-276, 2017.

[11] Jung, H., "The structural relationships between airline brand identification, brand image, brand relationship quality and brand loyalty," Korean Journal of Hospitality and Tourism, Vol. 30, No. 2, pp. 53-73, 2021.

[12] Kang, D. and Park, Y., "Fuzzy VIKOR method for evaluating service quality," The 2011 Fall Conference of Korean Institute of Industrial Engineers, pp. 743-748, 2011.

[13] Kang, S. and Kang, H., "Who Gets Government SME R&D Subsidy? Application of Gradient Boosting Model," The Journal of Society for e-Business Studies, Vol. 25, No. 4, pp. 77-109, 2020.

[14] Kim, S. and Kim, N., "A study on the effect of using sentiment lexicon in opinion classification," Journal of Intelligence and Information Systems, Vol. 20, No. 1, pp. 133-148, 2014.

[15] Korfiatis, N, Stamolampros, P., Kourouthanassis, P., and Sagiadinos, V., "Measuring service quality from unstructured data: A topic modeling application on airline passengers' online reviews," Expert Systems with Applications, Vol. 116, pp. 472-486, 2019.

[16] Leem, B., Cho, H., and Eum, S., "Using Sentiment Analysis of Passengers' Reviews to Analyze Airline Service Quality," Aviation Management Society of Korea, Vol. 18, No. 3, pp. 97-114, 2020.

[17] Lim, J. and Lee. H., "Comparisons of service quality perceptions between full service carriers and low cost carriers in airline travel," Current Issues in Tourism, Vol. 23, No. 10, pp. 1261-1276, 2020.

[18] Lucini, F. R., Tonetto, L. M., Fogliatto, F. S., and Anzanello, M. J., "Text mining

approach to explore dimensions of airline customer satisfaction using online customer reviews," Journal of Air Transport Management, Vol. 83, p. 101760, 2020.

[19] Machado, G., Mendoza, M. R., and Corbellini, L. G., "What variables are important in predicting bovine viral diarrhea virus? A random forest approach," Veterinary Research, Vol. 46, No. 1, pp. 1-15, 2015.

[20] Nadiri, H., Hussain, K., Ekiz, E. H., and Erdogan, S., "An investigation on the factors influencing passengers' loyalty in the North Cyprus national airline," The TQM Journal, Vol. 20, No. 3, pp. 265-280, 2008.

[21] Pakdil, F. and Aydın, Ö., "Expectations and perceptions in airline services: An analysis using weighted SERVQUAL scores," Journal of Air Transport Management, Vol. 13, No. 4, pp. 229-237, 2007.

[22] Parasuraman, A., Zeithaml, V. A., and Berry, L. L., "SERVQUAL: A multiple-item scale for measuring consumer perceptions of service quality," Journal of Retailing, Vol. 64, No. 1, pp. 12-40, 1988.

[23] Park, J. W., "Passenger perceptions of service quality: Korean and Australian case studies," Journal of Air Transport Management, Vol. 13, No. 4, pp. 238-242, 2007.

[24] Park, J. W., Robertson, R., and Wu, C. L., "The effect of airline service quality on passengers' behavioural intentions: a Korean case study." Journal of Air Transport Management, Vol. 10, No. 6, pp. 435-439, 2004.

[25] Park, S., Lee, J. S., and Nicolau, J. L., "Understanding the dynamics of the quality of airline service attributes: Satisfiers and dissatisfiers," Tourism Management, Vol. 81, p. 104163, 2020.

[26] Punel, A., Hassan, L. A. H., and Ermagun, A., "Variations in airline passenger expectation of service quality across the globe," Tourism Management, Vol. 75, pp. 491-508, 2019.

[27] Song, J. and Lee, S., "Automatic Construction of Positive/Negative Feature-Predicate Dictionary for Polarity Classification of Product Reviews," Journal of KISS: Software and Applications, Vol. 38, No. 3, pp. 157-168, 2011.

[28] Suleiman, D. and Awajan, A., "Comparative study of word embeddings models and their usage in Arabic language applications," In 2018 International Arab Conference on Information Technology (ACIT) (pp. 1-7). IEEE, 2018.

[29] Tsaur, S. H., Chang, T. Y., and Yen, C. H., "The evaluation of airline service quality by fuzzy MCDM," Tourism Management, Vol. 23, No. 2, pp. 107-115, 2002.

저 자 소 개

전우진 (E-mail: dnwls6314@ds.seoultech.ac.kr)
2020년 서울과학기술대학교 글로벌융합 산업공학과 (학사)
2020년～현재 서울과학기술대학교 데이터사이언스학과 석사과정
관심분야 Data Mining, Machine Learning, Natural Language Processing

이예빈 (E-mail: yebin@ds.seoultech.ac.kr)
2021년 서울과학기술대학교 글로벌융합 산업공학과 (학사)
2021년～현재 서울과학기술대학교 데이터사이언스학과 학석사연계과정
관심분야 Data Mining, Machine Learning, Natural Language Processing

금영정 (E-mail: yjgeum@seoultech.ac.kr)
2004년 KAIST 산업공학 (학사)
2013년 서울대학교 산업공학 (박사)
2014년～2020년 서울과학기술대학교 산업공학과 조교수
2020년～현재 서울과학기술대학교 산업공학과 부교수
관심분야 Technology Management, Data-driven innovation, Service Engineering, Business Model Innovation