



## 관광 빅데이터와 스몰데이터를 활용한 제주도 방문목적, 분위기, 인기토픽 비교연구

### A comparative study on the purposes of visiting, atmosphere, and popular topics of Jeju Island by using big and small tourism data

한학진\* · 서정모\*\* · 김재석\*\*\*

Han, Hagchin · Seo, Jung-mo · Kim, Jaeseok

\* 한남대학교 호텔항공경영학과 교수(주저자), hagchin@hnu.kr, 관심 분야: 관광정책, 관광자원

\*\* 제주한라대학교 호텔외식학과 교수(교신저자), sjm@chu.ac.kr, 관심 분야: 호텔경영, 외식경영

\*\*\* IATA 한국교육센터장, oxw@naver.com, 관심 분야: 관광정보시스템, 관광교육, 항공 서비스

#### 요약

4차 산업혁명의 시대가 도래함에 따라 인공지능을 비롯하여 사물인터넷, 자율 주행 서비스 등 기술이 고도화되고 있고, 개인과 기업 등 정보의 생성량이 급속도로 증가하면서 데이터의 양적인 확대와 질적인 활용으로 이어지고 있다. 본 연구는 코로나19와 포스트 코로나 기간 제주도의 방문목적, 분위기를 비롯하여 분위기 등 제주도의 전반적인 관광특성에 관한 관광객들의 행동과 인식을 관광 빅데이터와 스몰데이터를 활용하여 분석하고자 하였으며, 두 개의 분석결과를 항목별로 정리한 후 지역 맞춤형 시각화 기초분석 자료로 제공하고자 하였다. 관광 빅데이터와 스몰데이터의 분석결과에서 제주도 방문목적에서 '경치'와 '가족여행'은 일치하였고, 전반적인 만족도에 있어 관광 빅데이터(1차, 4.33)와 스몰데이터(2차, 4.36)간 거의 일치하고 있음을 확인하였다. 그럼에도 분위기와 인기토픽은 분석결과 간 전체적인 순위 및 구간 내 순위에서 크고 작은 차이가 나타났음을 확인하였다. 분석결과를 바탕으로 경치를 즐기는 가족여행을 위한 방문객이 가장 높게 나타남에 따라 관계기관에서도 난개발을 자제하여 현재와 같은 무공해의 자연환경을 지속가능하게 유지하는 것이 중요함을 제시하였다. 또한, 목표하고 있는 관광객에 중점을 둔 맞춤형 관광상품의 개발, 숙박형태, 식음료 부분의 신규 메뉴개발과 지역 상인의 친절서비스에 더욱 관심을 가져야 함을 강조하였다. 분석결과를 종합하여 제주도에 맞는 관광 안내지도 시각화 기초자료를 작성하여 제시함으로써 지역 내 관광정책 및 홍보마케팅에 적극적으로 활용할 수 있도록 제안하고자 하였다. 특히, 관광지 시각화의 주요편익으로 잠재 관광객들의 재방문과 추천이 가능하며 잠재 관광수요를 예측하고, 연구결과의 다른 지역으로의 확산 가능성을 제시하였다.

**핵심용어** 제주도, 방문목적, 분위기, 인기토픽, 관광 빅데이터, 스몰데이터

#### ABSTRACT

With the advent of the Fourth Industrial Revolution, technologies such as artificial intelligence, the Internet of Things, and autonomous driving services are becoming more advanced. The amount of information generated by individuals and companies is increasing rapidly, expanding the quantity and quality of data. This study aims to analyze tourists' behaviors and perceptions of Jeju Island's overall tourism characteristics, such as the purpose of visit and atmosphere during COVID-19 and post-COVID-19, by using tourism big data and small data, and the results of the two analyses were organized by item and provided as primary analysis data for regionally customized visualization. We found that 'scenery' and 'family travel' were consistent in visiting Jeju Island. There was a close match between the first (4.33) and second (4.36) rounds of tourism big data and small data regarding overall satisfaction. We also found significant and slight differences in the overall rankings and rankings within sections between the analyzed results. Based on the analysis results, it was emphasized that more attention should be paid to developing customized tourism products that focus on targeted tourists, accommodation types, new menus for tourists, and friendly services from local merchants. In this paper, the researcher created and presented the primary data for visualizing tourist guide maps tailored to Jeju Island to be actively used for regional tourism policies and promotional marketing. In particular, the main benefits of tourist destination visualization are that it allows potential tourists to revisit and make recommendations, predicts potential tourist demand, and suggests the possibility of spreading the research results to other regions.

**KEYWORDS** Jeju, visiting purposes, atmosphere, popular topics, tourism big data, small data

#### ACCEPTANCE INFO

원고접수 : 2023년 08월 16일 원고수정 : 2023년 09월 03일 게재확정 : 2023년 09월 11일 3인 익명심사필

#### ACKNOWLEDGEMENTS

본 연구는 2022년도 한남대학교 학술연구조성비 지원에 의하여 연구되었음.

## I. 서론

오늘날 첨단기술의 시대에서는 막대한 양의 다양하고 복잡한 데이터가 일상적으로 생산되며 이러한 데이터가 적극적으로 활용되고 있다. 데이터 사이언스의 발전은 빅데이터 개념의 형성으로 이어졌고, 실질적인 지식 생성에서 빅데이터 분석의 역할을 밝혀내며 풍부한 혁신적인 연구를 창출하고 있다(Kambatla et al., 2014). 4차 산업혁명의 시대가 도래함에 따라 인공지능기술을 비롯하여 사물인터넷, 자율주행 서비스에 분석기술이 고도화되고 있으며, 개인과 기업 등 정보의 생성량이 급속도로 증가하면서 데이터의 양적 확대와 질적 활용으로 이어지고 있다. 이는 고객과 조직 간의 관계에서 디지털 기술의 사용이 증가함에 따라 리뷰, 의견 텍스트 및 불만의 형태로 많은 양의 데이터를 제공하여 조직의 의사결정 개선에 필요한 지식을 추출할 수 있다. 인간은 인터넷과 소셜 네트워크를 통해 주제, 제품 및 서비스에 대한 좋은 의견이나 나쁜 의견을 공유하며, 다양한 종류의 정보를 효과적으로 분석하는 능력은 이제 의사결정을 더 잘 알리기 위한 핵심 경쟁 우위로 간주된다(El Alaoui & Gahi, 2019).

‘빅데이터’라는 용어와 관련하여 통계학자들은 크기, 많은 경우 또는 많은 변수 또는 둘 다의 측면에서 ‘큰’ 것을 생각하는 경향이 있으나 이 용어는 현상의 범위, 영향 및 마음 공유를 나타내는 ‘큰’ 과 함께 대중에게 더 넓은 의미를 지닌다. 빅데이터는 대규모 관찰 및 기계 분석을 처리하는 데 반해, 스몰데이터는 예측보다는 인과관계와 이해에 초점을 맞추는 휴먼 스케일의 실험적 또는 의도적으로 수집된 데이터의 결과로 이해된다(Lindstrom, 2016). 그동안 스몰데이터는 기존의 사회과학연구에서 연구대상이 적게는 수십 명에서 많게는 몇천 명 단위의 표본을 대상으로 분석한 결과를 제시하는 경우가 대부분이었다. 이처럼 빅데이터와 스몰데이터는 데이터의 양, 분석 처리의 복잡성, 처리, 비용 등의 측면에서 차이가 있음을 알 수 있다(Secchi, 2018).

한편, 관광 및 현대산업의 특정 분야를 조사하고 연구하는데 적용되는 새로운 접근 방식, 방법론 및 알고리즘이 지속적으로 등장하고 있다는 점이다(He et al., 2021; Ma et al., 2020). 예를 들어, 관광 빅데이터의

초기에는 텍스트마이닝, 연관어, 감성분석 등을 중심으로 진행된 연구가 연결망 분석을 비롯하여 기계학습, 딥 러닝의 형태로 고도화되고 있고, 최근에 와서는 Open AI ChatGPT, 바드, Bing과 같은 생성형 인공지능으로까지 확장되면서 인터넷 디지털 기반 개인 맞춤형 서비스로까지 확산일로에 있다.

정부는 지난 2019년을 시작으로 한국관광공사를 중심으로 관광 빅데이터 시범사업을 추진하고 있으며, 올해 5년 차에 접어들어 플랫폼의 고도화 시기에 접어들었다고 판단된다. 보다 구체적으로 살펴보면, 한국관광공사(2021)에서는 ‘빅데이터와 함께하는 똑똑한 컨설팅’ 사업은 전국 지자체 대상 공모사업으로 관광통계, 빅데이터(통신, 카드, 내비게이션, 소셜), 지자체 내부데이터(민원, 입장객 통계 등), 스몰데이터(지역주민·관광객 설문조사, 전문가 멘토링, 시민 아이디어 발굴단) 등을 활용해 무장애, 친환경, 지역상생, ESG(environmental, social and governance) 사업 등 지역맞춤형 사업을 도출하고 있다(한국관광공사, 2021).

지방공기업(예를 들어, 제주, 부산, 경북관광공사 등)과 관광 관련 연구원에서도 지역별로 빅데이터 전문조직과 인력을 확충하고 다양한 형태의 연구결과를 주기적으로 제시하고 있다. 제주관광공사(2021년)에서 관광 빅데이터(텍스트, 블로거, 유튜브 등)를 활용하여 분석한 제주관광 트렌드에 따르면, 코로나19 기간이 지속되어 해외여행이 어려워 지면서 국내 관광객이 다시 증가하고 있으며, 청정한 자연환경으로 인해 제주여행을 선택하고, 자연경관을 감상하거나 산/오름 등 트레킹 활동을 선호하며, 향후 제주에서 하고 싶은 활동으로 ‘자연경관 감상 > 식도락 > 트레킹’ 순으로 나타났음을 제시하였고, 코로나19 이후 캠핑, 오름 이외에 ‘차박’, ‘카라반’, ‘캠핑장/캠핑카’, ‘오토캠핑’에 대한 검색량이 증가하고 있음을 제시하였다(제주관광공사, 2021). 특히, 네이버를 비롯한 민간기업에서도 데이터랩(DataLab)을 활용한 관광지 분석자료를 오픈 소스(누구나 자유롭게 액세스할 수 있도록 설계된 코드)의 형태로 제시하고 있어 연구자들에게 빅데이터를 활용하여 연구, 분석의 기회를 제공하고 있다는 점은 매우 긍정적이라고 판단된다.

이에, 본 연구는 관광 빅데이터와 스몰데이터가 갖는 이러한 장단점 등의 본성에도 불구하고 국내 특정 지역을 대상으로 방문목적은 비롯하여 분위기, 그리

<표 1> 빅데이터와 스몰데이터의 개념 및 적용 범위 분석결과

구분	빅데이터	스몰데이터
용량	테라바이트, 페타바이트 또는 심지어 엑사바이트 단위로 측정되는 매우 큰 데이터 세트. 다양한 소스에서 수집된 방대한 양의 데이터가 포함	킬로바이트, 메가바이트 또는 기가바이트 단위로 측정되는 상대적으로 크기가 작은 데이터 세트. 단일 시스템에서 쉽게 관리
속도	실시간 또는 거의 실시간으로 고속으로 생성되고 수집. 다양한 소스에서 지속적으로 유입되는 데이터를 처리하려면 효율적인 처리가 필요	느린 속도로 수집되며 종종 배치 또는 주기적 업데이트로 수집. 빅데이터와 같은 고속 데이터 스트림을 포함하지 않음.
다양성	정형 데이터, 반정형 데이터, 비정형 데이터 등 다양한 데이터 유형을 포함. (텍스트 문서, 소셜 미디어 게시물, 이미지, 비디오, 센서 데이터 등)	더 구조화되고 단순. 종종 스프레드시트, 데이터베이스 또는 기타 조직화 된 형식에 저장된 구조화된 데이터로 구성. 설문조사 자료 포함
인프라 및 도구	NoSQL 데이터베이스 및 데이터 레이크와 같은 전문 인프라 및 기술이 필요. 복잡한 데이터 처리 파이프라인과 고급 분석기술 포함. 다양한 종류의 추가 비용이 발생하고 분석복잡	SQL(Structured Query Language) 쿼리, 스프레드시트, 통계 소프트웨어 패키지 및 시각화 도구가 포함. 더 빠르고 안정적, 저렴한 비용으로 올바른 결론에 도달
통찰력 및 의사결정	대규모 데이터 세트에서 패턴, 추세 및 상관관계를 밝히는 데 중점. 실행 가능한 통찰력을 도출하고, 예측하고, 데이터 기반 의사 결정을 지원	간단한 분석 방법을 사용하며 보다 관리하기 쉬운 데이터 세트에서 통찰력을 얻는 것을 목표. 핵심 성과 지표 추적, 고객 선호도 이해 또는 소규모 운영 모니터링과 같은 특정 목적

고 인기토픽 등을 비교 분석하고자 하였다. 이렇게 분석된 결과를 설문 조사와 같은 스몰데이터를 통해 확인하고자 하며, 보다 구체적으로는 플랫폼에서 제시한 기초분석결과에 관한 객관적인 측면의 정보 정확성을 확보하기 위해 비교분석이 필요하다고 판단되었다. 아울러 빅데이터를 활용한 연구는 커피전문점의 마케팅과 소비자의 인식변화(Wang Jie, 2023), 의료관광 인식 및 활성화 방안(조윤희, 2023), COVID-19 전후를 비교한 경제적 손실 추정(성보현 외, 2023) 등 연구가 진행되었으나, 스몰데이터와 관련된 연구는 스몰데이터의 유형에 따른 시각화 표현 방법(권지혜 외, 2019), Hadoop에서 스몰 데이터를 효율적으로 처리하기 위한 파일 가상 확장 기법(구본근, 2012), 그리고 빅데이터와 스몰데이터를 활용한 연구는 선형공원의 분석(심지수 & 오창송, 2020), 직무만족의 결정요인과 추천의도(이창재, 2022) 등 아직 연구가 미흡한 것을 확인할 수 있으며 특히, 관광분야 그리고 제주지역을 대상으로하는 빅데이터와 스몰데이터를 비교 연구한 논문은 거의 없다는 점을 인식하게 되었다.

본 연구의 보다 구체적인 목적은 다음과 같다. 첫째, 국내 플랫폼 기업의 데이터 랩의 오픈 소스에서 제시하고 있는 제주도 관광자원의 방문목적과 분위기, 인기토픽, 그리고 만족도를 제주방문 관광객을 대상으로 한 설문조사 형태인 스몰데이터 분석결과와 비교분석하고자 한다. 둘째, 두 개의 분석결과를 분야별로 종합적으로 정리한 후 이를 시각화 기초 작업을 통해 해당 지역에서 구체화할 수 있도록 제안하고자 한다. 데이터의 시각화가 중요한 이유는 의사 결정

자가 데이터 간의 관계를 식별하고 숨겨진 패턴이나 추세를 감지할 수 있고, 비즈니스 인텔리전스를 개선하고 데이터 중심 의사결정 및 전략적 계획 수립을 지원할 수 있다는 점이다. 이를 통해 해당 지역의 관광자원 개발 및 관리정책, 마케팅 및 홍보, 그리고 수요예측에 적극적으로 활용할 수 있을 것이다.

## II. 이론적 배경

### 1. 빅데이터와 스몰데이터의 개념 및 적용

빅데이터 분석은 데이터 속에 숨겨진 일정한 패턴을 비롯하여 규명되지 않은 변수 간의 관계 및 유용한 정보와 지식을 발견하기 위해 대량의 데이터를 수집, 분류하며 분석하는 프로세스를 지칭한다고 볼 수 있다(Singh. & Singh, 2012). 패턴, 정보 및 지식을 제공하여 비즈니스 이익을 높이고 기업경영의 효율성을 높이고 새로운 수요를 개척하거나 예측할 수 있는 장점으로 인해 주목을 받고 있다(LaValle et al., 2013). 기업이 더 나은 비즈니스 의사결정을 내리도록 지원하는 것을 주요 목표로 빅데이터 분석을 통해 사용하는 데이터베이스, 인터넷, 휴대폰 기록 및 위치, 센서 캡처 정보와 같은 다양한 소스에서 방대한 양의 데이터를 수시로 분석할 수 있다. 이러한 방대한 데이터를 다양한 형태로 분석하기 위해 빅데이터 분석과 관련된 기술은 광범위하게 적용되고 있다(Zakir et al., 2015).

<표 2> 국내 주요 선행연구 정리

연구자(연도)	주요 연구내용	비 고
박득희 외(2019)	제주도 방문 관광객들의 다 목적지 여행행동과 정보원천 간의 관계를 연결망 분석. 분석결과, 제주도에 위치한 다양한 목적지들 중 자연휴양형의 특징을 가진 목적지들이 핵심목적지로 나타남. 다음으로, 이원모드 중심 주변 모델 분석 결과 핵심블록은 8개의 목적지(한라산, 성산일출봉, 오설록, 용두암, 중문 관광단지, 서귀포, 섭지코지, 우도, 마라도, 추도, 테디베어 박물관)와 274명의 관광객으로 구성	네트워크 분석 (제주)
강영옥 외(2019)	텍스트마이닝과 딥러닝 기술을 활용한 외국인 관광객의 국내지역별 이미지 비교연구. 분석결과, "island", "Seogwipo", "spring", "city", "Hallasan"의 단어가 자주 출현하였고, 이미지 분석에서는 바다, 골짜기, 절벽 등의 사진을 많이 촬영함. 토픽모델링 결과 '자연경관', '도시경관', '쇼핑'이 상위토픽으로 분석. 이를 통해 제주도 방문 관광객의 주요 관심사가 자연경관(섬, 한라산, 바다, 골짜기, 절벽)임을 알 수 있고, 제주도는 독특한 지형적 특징을 가져 '자연경관'에 대한 관심이 높음	텍스트 마이닝 (제주)
고진현(2020)	한국 내의 일본제품 불매운동을 중심으로 사건 전후의 연관단어와 맥락, 시기별로 추이를 조사, 변화과정에 대한 이해와 시기별 빅데이터 네트워크 분석. 분석결과, 첫째, 불매운동의 본격화 시기인 2019년 하반기에는 '취소', '불매운동' 등의 연관단어들이 높은 빈도로 조사, 둘째, 연결 중심성 분석에서도 불매운동 기간에 '취소', '생각', '불매운동' 이 높은 연결 중심성을 보이면서 일본여행의 감소에 영향을 미침	네트워크 분석
주정 외(2021)	빅데이터를 이용하여 유산관광의 연관단어를 도출하고 소셜네트워크 분석. 분석결과, 유산관광 관련 연관단어의 빈도 분석결과 문화유산, 유네스코, 세계유산, 관광, 세계문화유산, 등재, 활용 순으로 빈도가 높게 나타났으며, 상위빈도 연관단어들의 주요 특성은 유네스코 등재유산, 문화관광 자원으로서의 활용, 역사 문화도시, 공적 정책기관 및 정책사업, 마케팅, 자원해설 관련 키워드들로 대별될 수 있었으며, 자연 및 복합유산에 대한 빈도는 매우 낮게 나타났음을 제시	연관어 분석
임종훈 & 하정우(2021)	서울 마포관광특구를 대상으로 소셜미디어 분석을 통하여 주목하는 콘텐츠와 핵심 자원 등을 파악. 분석결과 첫째, 축제, 거리, 관광객, 문화, 외국인, 주민 등 마포 흥대 문화를 대변하는 키워드들로 확인. 둘째, CONCOR 분석결과, 그 중 부동산, 합정동, 모델 하우스, 상권, 투자, 호텔, 아파트, 분양, 오피스텔의 집단을 보면, 기초수립 단계에 있어 주민들의 우려 대상인 임대료 상승 부분을 확인할 수 있는 키워드들로 구성	소셜미디어 분석
이창재(2022)	온라인 리뷰에 기반한 직무 만족의 결정요인과 추천의도 연구를 빅데이터 분석과 스몰데이터 활용 비교분석. 취업 포털 사이트인 잡플래닛에서 9개 업종별로 대량의 데이터를 수집하고 비교·분석하여 기존의 구조화된 설문문의 한계점을 보완하고 유의미한 결과를 도출함. 또한, 스몰데이터와 빅데이터 분석결과가 큰 차이가 없고, 각 방법이 가지고 있는 장점과 한계점을 상호 보완하면 기업의 인사운영 계획 수립을 위한 의사결정 도구로 유용하게 사용될 수 있음을 제시	빅데이터와 스몰데이터 비교

출 처: Open AI ChatGPT & Faraway(2018)의 자료를 중심으로 논자 재구성

데이터에는 획득, 계산 및 개인 정보 보호 비용이 포함되어 비용과 이점의 균형을 맞춰야 하는데, 통계적 추론은 작은 데이터에서는 잘 작동하지만 큰 데이터에서는 잘 작동하지 않는 경우가 있다. 빅데이터는 대규모 관찰 및 기계 분석을 통해 처리하며 작은 데이터는 예측보다는 인과관계와 이해에 초점을 맞추는 휴먼 스케일의 실험적 또는 의도적으로 수집된 데이터의 결과로 판단된다(Faraway, 2018). 다음은 빅데이터와 스몰데이터의 차이에 관해 정리하였다. 이를 보다 구체적으로 살펴보면, <표 1>에 나타난 바와 같다. 이같이 빅데이터와 스몰데이터의 구분이 항상 엄격한 것은 아니라는 점에 유의하고, 정의 및 값은 콘텍스트, 산업 및 사용 가능한 원자료에 따라 다를 수 있어 주요한 차이점은 데이터 관리 및 분석과 관련된 규모, 복잡성 및 인프라 요구 사항에 있다고 판단된다.

## 2. 빅데이터의 관광 분야 연구에의 적용 및 이슈

관광 분야에서 빅데이터를 활용한 연구의 질적 및 양적 증가는 관광지를 비롯하여 호텔산업의 수익성 분석 및 평가관리에 유용하면서 실용적인 지식을 제공하고 있다. 그럼에도, 빅데이터 연구는 방법론이 다소 제한적이며 의미 있는 지식 창출을 위해서는 보다 견고한 이론적, 철학적 기반이 필요함을 강조하였다(Kambatla et al., 2014). 관광 분야의 데이터 트렌드와 방법론적 접근, 그리고 연구 주제를 일목요연하게 보여주고 있는 Lyu et al.(2022)의 연구는 관광 빅데이터를 활용한 의미 있는 연구로서 분석에 활용된 원천자료, 방법론 및 분석기법, 그리고 연구초점을 중심으로 제시하고자 한다. 첫째, 원천자료와 관련하여 관광 및 숙박업 연구에 사용되는 빅데이터는 사용자 제작 콘텐츠(UGC)(72%), 운영(웹 검색)(17%), 사물인터넷

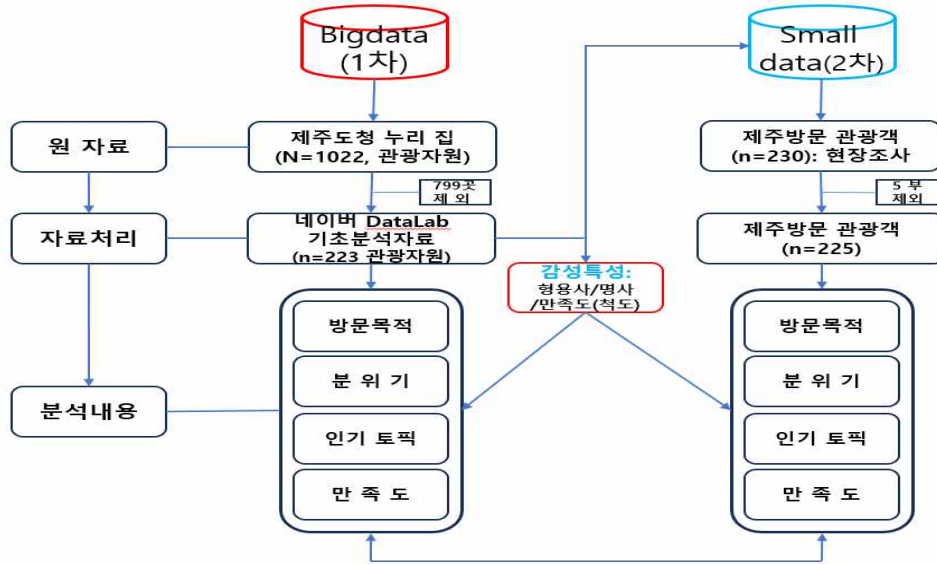
(IoT)(10%)의 세 가지 주요 소스에 기반하고 있음을 알 수 있다. 둘째, 데이터의 특성과 연구 목적에 따라 관광 및 호텔 연구 분야에서 빅데이터를 분석하기 위해 다양한 방법론 및 분석기법이 사용되었는데, 처음 10년 동안은 주로 기술 통계, 콘텐츠 분석, 빈도 분포 및 상관관계에 중점을 두었다면, 최근에는 공간, 소셜 네트워크, 인공 신경망 분석 및 기계학습 모델링 접근 방식과 같은 고급 분석 방법을 적용하고 있다는 점이다(Giglio et al., 2019; Guo et al., 2017). 셋째, 연구초점은 주로 관광객의 심리, 관광 및接客 분야의 빅데이터 및 인공지능 및 기타 관련 비즈니스에 초점을 맞추고 있으며, 호텔 점유율, 관광 패턴, 관광명소, 목적지 이미지, 관광객 만족도, 인식 등을 반복적으로 연구되고 있다. 보다 구체적으로는, 온라인 리뷰, 여행자의 블로그, 트윗 및 일기에서 오는 것으로 관찰되며, 온라인 리뷰가 일반적으로 관광객 행동(지출 및 소비 패턴 및 경험), 호텔 서비스에 대한 태도, 관광명소, 목적지에 대한 인식, 서비스 표준에 대한 선호도 및 만족도 속성과 같은 심리적 측면에 초점을 맞추고 있음을 알 수 있다(El Alaoui & Gahi, 2019; Gong et al., 2016). 또한, 온라인 텍스트 데이터와 유사하게 사진 데이터도 관광객 행동, 이동 패턴, 경험, 디지털 발자국 및 목적지 이미지를 분석하는 데 널리 사용되었고, GPS, 휴대폰, 블루투스 등 관광 및接客 분야에서 자료수집을 위해 다양한 장치가 활용되고 있다.

한편, 빅데이터 샘플이 통계적 유의성 개념과 양립되지는 않으나 스몰데이터 통계 기법 및 방법은 통계적 유의성을 기반으로 한다. 스몰데이터는 빅데이터와 상대적인 관점에서 간주해야 할 것으로 판단된다. 기존의 사회과학연구에서 연구대상이 적게는 수십 명에서 많게는 몇천 명 단위의 표본을 대상으로 분석한 결과를 제시하는 경우가 대부분이었다. 다시 말해, 질적 연구와 양적 연구에 스몰데이터가 적용되고 이를 연구에 활용하는 연구들이 있다고 판단된다. 그동안 여행 및 관광 분야의 빅데이터는 신뢰성 문제에 직면해 있는데, 예를 들어 기만적인 리뷰의 가능성이 높다(Martinez-Torres & Toral, 2019). 따라서 가짜 리뷰를 탐지할 수 있는 메커니즘을 개발하고, 통계처리 및 방법론적 프레임워크 개발은 연구자에게 상당한 도전이 될 것임을 제시하였다(Liu & Ji, 2018).

### 3. 빅데이터 관련 국내외 선행연구

국내외 선행연구에 관해 관광 빅데이터와 관련된 다수의 연구가 진행된 것은 사실이다. 우선, 연구의 내용적 및 방법론적 측면에서 접근이 가능하다. 이를 보다 구체적으로 살펴보면, <표 2>에 나타난 바와 같다.

한편, 관광 빅데이터에 관한 해외 선행연구는 다음과 같다. 이를 보다 구체적으로 살펴보면, i Agusti(2021)는 다양한 연구에서 성별에 따른 관광객의 공간적 행동 차이를 연구하였다. Instagram을 사용하여 관광객이 촬영하는 장소 및 주제와 관련된 다양한 성별 관련 형태의 관광객 행동을 식별할 수 있도록 하였다. 연구결과, 인스타그램에서 관광객이 공유하는 명소의 공간적 분포에서 성별 차이가 나타났으며, 관광객이 혼자 여행하는지 동행하는지 여부가 공간 행동에 영향을 미치는 것으로 나타났음을 제시하였다. 본 연구의 경영상의 함축적 의미는 인스타그램 이미지에서 얻은 결과는 다른 소셜 네트워크에 설명된 것과 일치했으며 인터뷰와 같은 다른 수단을 통해 얻은 결과와도 일치했음을 제시하였다. Garner et al.(2022) 외는 여행 차원에 대한 <Yelp.com> 리뷰의 의견을 해석하고 분류하기 위해 토픽모델링을 도입하였다. 연구결과, 자연이 소비자 행복에 미치는 데, 자연 자체는 전통적인 제품 의미에서 ‘소모품’ 이 아니라 감상과 경험적 참여를 수반한다는 점과 지역 문화와 장소 애착이 행복 수준의 강력한 지표라는 점을 제시함으로써 여행 경험은 독특하기 때문에 더 큰 소비자 행복을 유도할 수 있다는 점을 강조하였다. Jardim & Mora(2022)는 게시물에 주관적으로 표현된 감성표현의 극성(polar)에 따라 등록된 리뷰나 댓글에서 사용자가 표현한 감성을 추출하고 이에 따른 사용자의 자동 클러스터링을 통해 디지털 관광 플랫폼 사용자 간의 유사점을 클러스터링하고 식별할 수 있는 방법을 제안하였다. 연구결과, 제안된 솔루션의 아키텍처는 첫 번째는 분석 및 분류할 텍스트의 전처리를 위한 것이고, 두 번째는 텍스트가 작성된 언어에 해당하는 사전을 사용하여 감정 식별 및 추출을 위한 것이며, 세 번째는 식별된 정서의 분류 및 표현된 정서의 글로벌 극성 계산, 네 번째는 게시물의 정서 극성을 기반으로 사용자 클러스터링을 위한 것이며, 솔루션은 4개 언어로 작성된 사용자 게시물 분석에서 테스트 되었으며 감정 분류 및 사용자 세분화에 대한



<그림 1> 연구모형

높은 정확도를 보여주고 있음을 확인하였다. Lyu et al.,(2022)은 빅데이터 철학에 대한 관광전공 학자들에게 중요한 안목을 제공하고, 데이터 유형, 방법 및 연구초점을 연결하는 프레임워크를 개발하며, 관광 관리 관행을 안내하는 빅데이터 연구의 역할을 밝히고자 하였다. 146개의 빅데이터 관련 기사에 대한 내용 분석을 통해 첫째, 목적지, 호텔, 수익 및 평판 관리에 대한 유용하고 실용적인 지식을 제공하였으며, 둘째, 관광 및 호텔 분야의 빅데이터 연구가 목적지 관리, 이미지 구축, 기후 변화 및 호텔 성과, 수익 관리, 경험 관리, 공급망 관리 및 평판 관리에 실질적인 영향을 미친다는 것을 확인시켜줌으로써 가치 있는 지식을 개발하기 위해 데이터 수집 및 분석기술의 실험적 및 개념적 설계를 안내하는 이론을 활용하는 것이 필수적임을 제시하였다.

이처럼 국내외의 다양한 선행연구에서 볼 수 있듯이 소셜 미디어 콘텐츠와 감정 분석 방법을 기반으로 사용자를 추출하고 측정하며, 특정 주제에 대한 피드백과 실제로 감성분석(SA: sentiment analysis)은 소셜 콘텐츠에서 놀라운 숨겨진 연결을 발견할 수 있게 해주는 빅데이터에서 매력적인 연구 분야가 되고 있다(Gong et al., 2016). 그럼에도 관광 빅데이터와 스몰데이터를 활용하여 특정 관광지의 관광특성에 관한 관광객의 인식을 비교하여 진행한 연구는 미흡함을 알 수 있다.

### III. 연구 설계

#### 1. 연구모형

본 연구를 진행하기 위한 프로세스는 <그림 1>과 같다.

#### 2. 연구과제

가. 관광 빅데이터와 스몰데이터 간 제주도 방문목적에 따른 전체 및 구간 내에서의 순위는 차이가 있을 것인가?

나. 관광 빅데이터와 스몰데이터 간 제주도 분위기에 따른 전체 및 구간 내에서의 순위는 차이가 있을 것인가?

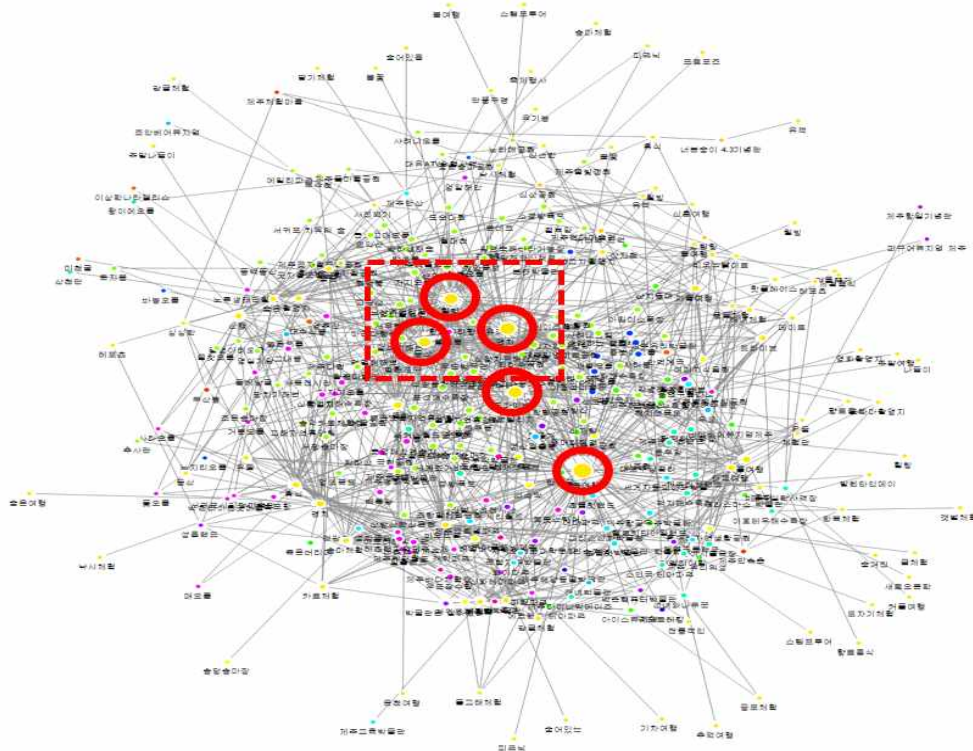
다. 관광 빅데이터와 스몰데이터 간 제주도 인기토픽에 따른 전체 및 구간 내에서의 순위는 차이가 있을 것인가?

라. 관광 빅데이터와 스몰데이터 간 제주도의 전반적인 만족도는 유의적인 차이가 있을 것인가?

#### 3. 자료수집 및 조사설계

빅데이터와 관련 기술의 등장은 감성분석 영역에 큰 기회를 제공하고 있으며, 그러나 데이터 분석을 완전히 방해할 수 있는 빅데이터 품질과 같은 데이터와 관련된 다소 중요한 이슈가 있는 것 또한 사실이다. 감성분석(SA)은 고객 리뷰, 의견 텍스트 및 평가에서 주관적으로 표현된 감정을 분석하는 기법으로 소비자





<그림 2> 방문목적에 관한 연결망 분석결과

의 구매 패턴 분석, 소셜 미디어, 웹사이트 또는 온라인에서 고객의 피드백을 수집하고, 사람들이 제품이나 서비스를 좋아하도록 동기를 부여하는 요인을 이해하며, 고객의 서비스 요청 분류를 통한 소비자 행동을 이해하고 예측하는데, 활용되고 있다(Jardim & Mora, 2022).

본 연구에서는 관광 빅데이터와 스몰데이터를 활용하여 제주관광에 대한 일반인들의 인식을 분석하기 위해 국내 대표 포털 사이트인 네이버의 데이터 랩(Datalab)의 오픈 소스와 제주 방문객 설문조사를 병행하여 활용하였다. 제주도에 관한 빅데이터 자료의 수집 기간은 2021년 9월부터 12월까지(4개월간)였다. 수집내용은 제주 관광자원 전체를 확인하였고, 제주도청 누리집에 게시된 관광자원 전체는 1,022건이었다. 수집절차는 우선, 제주도 누리집 방문, 게시되고 분류된 관광지 현황을 확인한 후 네이버 데이터 랩의 오픈 소스에서 분석 제공된 내용이 있는 관광자원을 확인하여 추출 및 분류하였고, 이를 엑셀자료로 정리하였다. 포함된 내용은 방문목적, 분위기, 그리고 인기토픽, 만족도, 그리고 성별 연령별 검색순위였다. 제주도청의 누리집에 게시된 전체 관광지는 1,022건이었는데, 그중 네이버 데이터 랩에서 기초적으로 제

시된 관광지는 전체 233곳이었다. 이렇게 수집된 자료를 지역별(제주시와 서귀포시), 관광지 유형별(자연 및 인문 관광자원)로 분류하고 정제하는 과정을 거쳤다. 정제과정에서 비슷하거나 동의의 뜻을 가진 용어(예, 이국적인 혹은 이국적)는 한 가지로 통일하여 정리하였다. 이렇게 정리된 원자료를 관광 빅데이터의 SNA(사회연결망) 분석기법을 적용하였다. 분석내용은 각 키워드 간의 연결성과 중심성을 분석하였고, 아이겐벡터(eigenvector)값을 통한 중요도 순위를 제시하였다. 분석된 내용에서 방문목적은 36개, 분위기는 36개, 그리고 방문목적은 80개가 제시되었다. 한편, 1차적으로 관광 빅데이터를 활용하여 분석된 내용에 기반하여 스몰데이터를 활용한 설문을 완성하였다. 최종설문 문항에서는 1차 분석결과에서 제시한 방문목적, 분위기, 방문목적 외에 제주도 방문횟수, 전반적인 만족도, 그리고, 성별과 연령 등 인구통계적 사항 2개 문항으로 구성하였다. 수정 및 보완된 설문을 최종적으로 확정하였고, 확정된 설문을 제주도를 방문한 관광객을 대상으로 설문을 진행하였다. 설문은 제주공항에서 2022년 9월부터 12월까지 진행하였으며 설문에 응한 대상은 230명이었고, 최종 유효분석에는 225명을 사용하였다. 관광 빅데이터 분석(1차)

<표 3> 방문목적에 관한 빅데이터와 스몰데이터 간 분석결과

구간	순 위	문 항	1차 분석	순 위	문 항	2차 분석	
			(n = 223) eigenvector			(다중응답, n = 1126) N %	
1	1	경치	0.5027	1	경치	85	7.5
	2	가족여행	0.4834	2	가족여행	83	7.4
	3	드라이브	0.3635	3	데이트	69	6.1
	4	힐링	0.3209	4	드라이브	66	5.9
	5	등산	0.2629	5	사진찍기	63	5.6
2	6	비오는 날	0.2243	5	핫플레이스	63	5.6
	7	나들이	0.1585	7	힐링	58	5.2
	8	숨은여행	0.1585	8	등산	56	5.0
	9	휴 식	0.1469	9	향토음식	51	4.5
	10	커플여행	0.1160	10	휴 식	44	3.9
3	11	체험관	0.1082	11	낚시체험	38	3.4
	12	산행	0.1044	12	감귤체험	36	3.2
	13	촬영지	0.1044	12	산행	36	3.2
	14	승마체험	0.0928	14	레포츠	35	3.1
	15	태교여행	0.0812	15	숨은장소	33	2.9
4	16	가을여행	0.0734	16	촬영지	31	2.8
	17	데이트	0.0657	17	커플여행	24	2.1
	18	유물	0.0618	18	승마체험	22	2.0
	19	감귤체험	0.0541	19	가을여행	21	1.9
	20	카트체험	0.0348	19	봄꽃여행	21	1.9
5	21	봄여행	0.0309	21	신혼여행	18	1.6
	22	사진찍기	0.0309	22	웰빙	16	1.4
	23	동물체험	0.0232	23	체험관	15	1.3
	24	해녀체험	0.0232	23	가을축제	15	1.3
	25	낚시체험	0.0193	25	동물체험	9	0.8

에 사용된 최종자료는 233개의 제주도 관광자원이 대상이었고, 스몰데이터(2차) 분석에 사용된 최종자료는 제주도를 방문한 관광객(n=225)이었다. 분석된 세부내용은 방문목적, 관광 분위기, 인기토픽, 전반적인 만족도, 방문횟수, 그리고 인구 통계적 사항 등이 포함되었다.

#### IV. 실증 분석

##### 1. 분석결과

###### 1) 인구통계적 사항 및 여행특성 분석결과

인구통계적 특성에서 전체 설문대상자 225명 중 성별과 연령별 분석결과는 성별에서는 남성이 109명으로 전체 48.4%를 차지하였고, 여성은 116명으로 51.6%로 나타났다. 연령대에서는 30대가 59명으로 26.3%를 차지하였고, 40대가 47명으로 21%로 나타났으며, 20대 18.8%(42명), 50대 17%(38명), 60대 이상이 9.8%(22명)의 순으로 나타났다. 다음으로, 제주도 방문횟수와 관련한 분석결과, 제주도 방문횟수가 4-6회

가 69명으로 나타나 전체 31%를 차지하였고, 7-9회가 22%로 나타나 그다음을 차지하였으며, 13회 이상이 19%, 10~12회가 14%로 나타났으며, 1~3회는 13.8%로 가장 낮은 비율로 나타나고 있음을 알 수 있다.

###### 2) 방문목적에 관한 분석결과

다음은 제주도의 방문목적에 관한 관광 빅데이터 분석결과이다. 이를 보다 구체적으로 살펴보면, <그림 2>에 제시된 바와 같다.

###### ① 방문목적에 관한 연결망 분석결과(1차)

<그림 2>에 제시된 바와 같이, 방문목적에 관한 연결 중심성을 분석한 결과, 전체적인 큰 그림에서 그룹 1은 ‘경치’를 중심으로 하여 ‘힐링’과 ‘등산’이 근접하게 위치하고 있고, 반대편에는 ‘가족여행’을 중심으로 한 다양한 노드들이 존재한다. 특히, ‘경치’, ‘힐링’, ‘등산’과 ‘가족여행’ 중심에는 ‘드라이브’가 중앙에 위치하여 매개 역할을 하거나, 각 요소에 중요한 역할을 하고 있음을 확인하였다. 이러한 제주도의 방문목적 특성은 연결망으로 볼 때, 경치를 중심으로 한 관광과 가족여행을 중심으로



한 여행이 크게 구분됨을 확인할 수 있었다. 특히, 코로나19 시기 및 포스트 코로나 시기에 가족 단위의 관광객들이 제주도의 경치를 보기 위해 방문했음을 알 수 있었다.

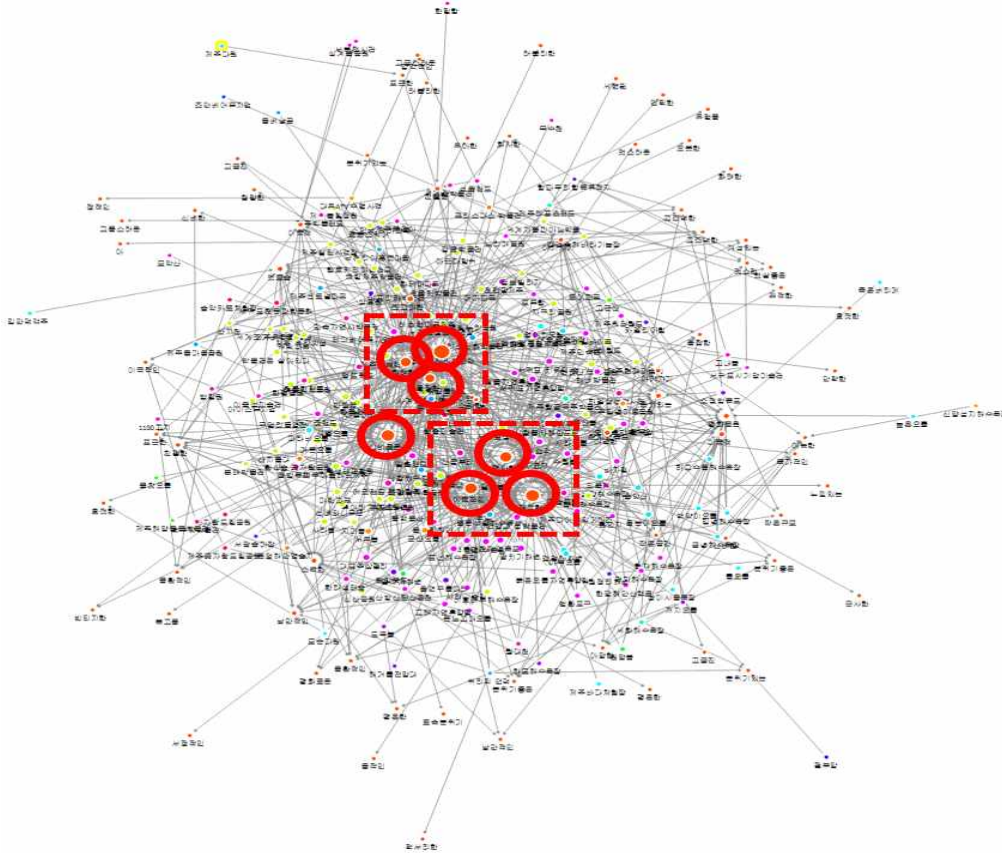
② 방문목적에 관한 관광 빅데이터(1차)와 스몰데이터(2차) 비교분석

다음은 방문목적에 관해 관광 빅데이터와 스몰데이터를 활용하여 분석하였다. 이를 보다 구체적으로 살펴보면, <표 3>에 제시된 바와 같다.

위), 가족여행(2위), 데이트(3위), 드라이브(4위), 사진찍기(5위)의 순으로 나타났다. 이를 통해 제주도 방문 목적은 ‘경치’와 ‘가족여행’의 순으로 나타났고, 두 개의 분석결과와 순위에서 일치하고 있음을 알 수 있다. 특히, 1구간에서의 순위의 일치 개수가 3개로 나타났다.

3) 분위기에 관한 분석결과

① 분위기에 관한 연결망 분석결과(1차)



<그림 3> 분위기에 관한 연결망 분석결과

<표 3>에 나타난 바와 같이, 여기에서 고유벡터(eigenvector)란 어떤 선형변환을 취했을 때, 방향은 변하지 않고 크기만 변하는 벡터를 의미한다. 빅데이터의 고유벡터값과 스몰데이터의 빈도 혹은 구성비(%)를 절대 값으로 비교하기가 어려워 순위로서 비교하였으며, 순위의 분석결과를 5개씩을 한 구간으로 설정하여 제시하고 이들 전체순위 및 구간 내 순위를 비교하고자 한다. 관광 빅데이터 분석(1차)에서 경치(1위, 0.5027), 가족여행(2위, 0.4834), 드라이브(3위, 0.3635), 힐링(4위, 0.3209), 등산(5위, 0.2629)의 순으로 나타났다. 한편, 스몰데이터 분석(2차)에서는 경치(1

다음은 제주도의 분위기에 관한 관광 빅데이터 분석결과이다. 이를 보다 구체적으로 살펴보면, <그림 3>에 제시된 바와 같다.

<표 3>에 나타난 바와 같이, 여기에서 고유벡터(eigenvector)란 어떤 선형변환을 취했을 때, 방향은 변하지 않고 크기만 변하는 벡터를 의미한다. 빅데이터의 고유벡터값과 스몰데이터의 빈도 혹은 구성비(%)를 절대 값으로 비교하기가 어려워 순위로서 비교하였으며, 순위의 분석결과를 5개씩을 한 구간으로 설정하여 제시하고 이들 전체순위 및 구간 내 순위를 비교하고자 한다. 관광 빅데이터 분석(1차)에서 경치(1

<표 4> 분위기에 관한 빅데이터와 스몰데이터 간 분석결과

구간	순위	문항	1차 분석	순위	문항	2차 분석	
			(n=223) eigenvector			(다중응답, n=1128) N %	
1	1	재미있는	0.4624	1	평온한	93	8.2
	2	이국적	0.4535	2	깨끗한	77	6.8
	3	깨끗한	0.3824	3	분위기 좋은	76	6.7
	4	신비로운	0.3557	4	편안한	61	5.4
	5	편안한	0.2935	5	쾌적한	55	4.9
2	6	화려한	0.2846	6	재미있는	50	4.4
	7	웅장한	0.2179	6	평화로운	50	4.4
	8	아기자기	0.1067	8	작은공간	48	4.3
	9	아담한	0.1067	9	아담한	43	3.8
	10	조그마한	0.1067	10	아기자기	40	3.5
3	11	소박한	0.1022	11	낭만적인	38	3.4
	12	친절한	0.0978	11	조그마한	38	3.4
	13	평화로운	0.0978	13	신비로운	36	3.2
	14	운치있는	0.0755	13	옛모습	36	3.2
	15	옛모습	0.0711	13	친절한	36	3.2
4	16	분위기 좋은	0.0667	13	소박한	36	3.2
	17	낭만적인	0.0533	17	개성있는	29	2.6
	18	고즈넉한	0.0444	18	멋스런	28	2.5
	19	몽환적인	0.0444	18	포근한	28	2.5
	20	작은규모	0.0355	20	운치있는	27	2.4
5	21	쾌적한	0.0311	21	고급스러운	25	2.2
	22	포근한	0.0311	22	이국적인	24	2.1
	23	고급진	0.0266	22	우아한	24	2.1
	24	아늑한	0.0266	24	아늑한	23	2.0
	25	화사한	0.0177	25	웅장한	18	1.6

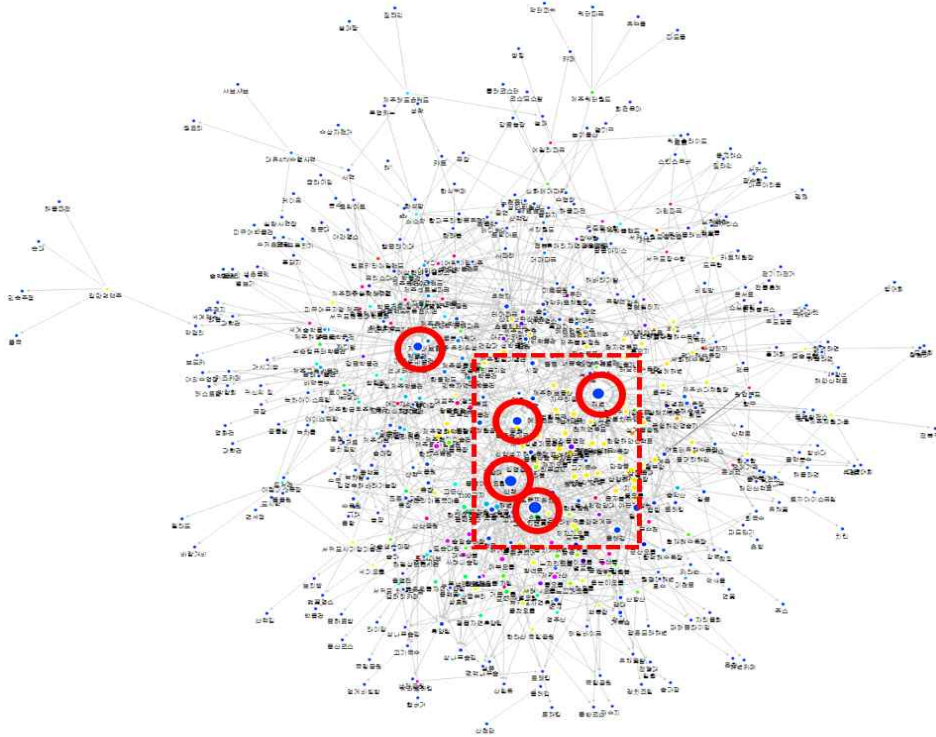
위, 0.5027), 가족여행(2위, 0.4834), 드라이브(3위, 0.3635), 힐링(4위, 0.3209), 등산(5위, 0.2629)의 순으로 나타났다. 한편, 스몰데이터 분석(2차)에서는 경치(1위), 가족여행(2위), 데이트(3위), 드라이브(4위), 사진 찍기(5위)의 순으로 나타났다. 이를 통해 제주도 방문 목적은 ‘경치’와 ‘가족여행’의 순으로 나타났고, 두 개의 분석결과의 순위에서 일치하고 있음을 알 수 있다. 특히, 1구간에서의 순위의 일치 개수가 3개로 나타났다.

<그림 3>에 제시된 바와 같이, 분위기에 관한 연결망을 분석한 결과, 전체적인 큰 그림에서 그룹 1은 ‘재미있는’, ‘웅장한’, ‘화려한’이 위치하고 있으며, 그룹 2는 ‘이국적’, ‘깨끗한’, ‘편안한’이 위치하고 있고, ‘신비로운’이 별도로 위치하고 있음을 확인할 수 있었다. 이처럼 제주도의 분위기 특성을 살펴보면, 화려하고 웅장하며 재미있는 요소를 중심으로 한 관광과 이국적이며 깨끗하고 편안한 분위기의 요소로 크게 구분됨을 확인하였다.

② 분위기에 관한 관광 빅데이터(1차)와 스몰데이터(2차) 비교분석

다음은 분위기에 관한 관광 빅데이터와 스몰데이터를 활용하여 분석하였다. 이를 보다 구체적으로 살펴보면, <표 4>에 제시된 바와 같다.

<표 4>에 나타난 바와 같이, 관광 빅데이터 분석(1차)에서 ‘재미있는’ (1위, 0.4624), ‘이국적’ (2위, 0.4535), ‘깨끗한’ (3위, 0.3824), ‘신비로운’ (4위, 0.3557), ‘편안한’ (5위, 0.2935)의 순으로 나타났다. 한편, 스몰데이터 분석(2차)에서는 ‘평온한’ (1위), ‘깨끗한’ (2위), ‘분위기 좋은’ (3위), ‘편안한’ (4위), ‘쾌적한’ (5위)의 순으로 나타났다. 이를 통해 제주도 분위기는 1차 분석에서는 ‘재미있고’, ‘이국적이며’, ‘깨끗한’으로 분석된 반면, 2차 분석에서는 ‘평온한’, ‘깨끗한’, ‘분위기 좋은’으로 나타나 두 개의 분석결과의 순위에서 일치하지 않았다. 종합하면, 1구간에서 분위기가 ‘깨끗하다’, ‘편안하다’가 포함되었음을 알 수 있고, 나머지 구간에서의 순위에는 차이가 나타났음을 알 수 있다.



<그림 4> 인기토픽에 관한 연결망 분석결과

3) 인기토픽에 관한 분석

① 인기토픽에 관한 연결망 분석결과(1차)

다음은 제주도의 인기토픽에 관한 관광 빅데이터 분석결과이다. 이를 보다 구체적으로 살펴보면, <그림 4>에 제시된 바와 같다.

<그림 4>에 제시된 바와 같이, 인기토픽 중심성에 관한 연결망 분석결과, 전체적인 큰 그림에서 그룹에서 ‘오름’, ‘해변’, ‘산책’, ‘공원’ 중심으로 하고, 그 중심에서 약간 떨어진 위치에 ‘박물관’ 이 높은 중심성을 보였다. 이러한 결과는 제주도의 인기토픽이 ‘오름’, ‘해변’, ‘산책’, ‘공원’, ‘일몰’, ‘둘레길’ 과 같은 주변 환경을 중심으로 한 풍경이 높은 중심성을 가지고 있음을 확인하였으며, 또한, ‘박물관’, ‘테마파크’ 와 같은 관광시설 등도 인기토픽의 키워드로 중요성을 가지고 있음을 확인하였다.

② 인기토픽에 관한 관광 빅데이터(1차)와 스몰데이터(2차) 비교분석

다음은 인기토픽에 관해 관광 빅데이터와 스몰데이터를 활용하여 분석을 실시하였다. 이를 보다 구체적으로 살펴보면, <표 5>에 제시된 바와 같다.

<표 5>에 나타난 바와 같이, 관광 빅데이터 분석(1차)에서 ‘오름’ (1위, 0.4769), ‘해변’ (2위, 0.4137),

‘박물관’ (3위, 0.3620), ‘산책’ (4위, 0.3217), ‘공원’ (5위, 0.3045)의 순으로 나타났다. 한편, 스몰데이터 분석(2차)에서는 ‘고기국수’ (1위), ‘갈치’ (2위), ‘오메기떡’ (3위), ‘해변’ (4위), ‘흑돼지’ (5위)의 순으로 나타났다. 이를 통해 제주도 인기토픽은 1차 분석에서는 ‘오름’, ‘해변’, ‘박물관’ 등으로 분석된 반면, 2차 분석에서는 ‘고기국수’, ‘갈치’, ‘오메기떡’ 으로 나타나 1구간에서의 두 분석 결과의 순위에서는 해변만이 일치하고 나머지는 일치하지 않았음을 알 수 있다. 종합하면, 제주도의 인기토픽에서는 ‘해변’ 을 제외하고 거의 모든 구간 및 구간 내 순위에서 일치하지 않은 것으로 분석되었다. 특히, 먹방과 관련하여 음식 관련 키워드가 두드러지게 순위에 올랐다는 점이다. 이러한 결과는 인기토픽은 방문목적이나 분위기와 달리 단시간에 인기의 흐름 혹은 트렌드가 변화할 수 있다는 점을 나타내고 있음을 알 수 있다.

4) 전반적인 만족도 비교분석

다음은 제주도의 전반적인 만족도를 빅데이터와 스몰데이터를 활용하여 분석하였다. 관광 빅데이터 분석(1차) 결과, 제주도내 223곳의 관광자원(관광지 포함)의 전반적인 만족도는 평균 4.33(리커트 5점 척도)

<표 5> 인기토픽에 관한 빅데이터와 스몰데이터 간 분석결과

구간	순위	문항	1차 분석	순 위	문 항	2차 분석	
			(n = 223) eigenvector			(다중응답, n = 862)	N
1	1	오름	0.4769	1	고기국수	103	11.9
	2	해변	0.4137	2	갈치	95	11.0
	3	박물관	0.3620	3	오메기 떡	84	9.7
	4	산책	0.3217	4	해변	74	8.6
	5	공원	0.3045	5	흑돼지	65	7.5
2	6	일몰	0.1953	6	밤바다(야경)	43	5.0
	7	둘레길	0.1436	6	동백꽃	43	5.0
	8	전망대	0.1321	8	오름	33	3.8
	9	데마파크	0.1321	9	해물라면	32	3.7
	10	폭포	0.1264	10	감귤농장	28	3.2
3	11	동굴	0.1149	11	녹차밭	27	3.1
	12	유채꽃	0.1149	12	둘레길	24	2.8
	13	역새	0.0976	13	숲길	21	2.4
	14	목장	0.0919	13	드라이브길	21	2.4
	15	시장	0.0919	15	등반코스	20	2.3
4	16	공연	0.0861	15	산책	20	2.3
	17	야경	0.0861	17	디저트	19	2.2
	18	계곡	0.0804	18	카드	17	2.0
	19	산책로	0.0747	19	횃집	15	1.7
	20	캠핑	0.0747	20	일몰	12	1.4
5	21	승마	0.0689	20	캠핑	12	1.4
	22	식물원	0.0632	22	박물관	11	1.3
	23	고기국수	0.0574	22	미로공원	11	1.3
	24	야간명소	0.0574	24	유채꽃	10	1.2
	25	흑돼지	0.0517	24	놀이동산	10	1.2

으로 나타났다. 다음으로, 스몰데이터 분석(2차)을 통해 나타난 전반적인 만족도는 매우만족은 104명으로 46.2%로 나타났고, 만족은 99명으로 44%, 보통은 21명으로 9.3%로 나타나 제주를 방문한 방문객의 만족도는 평균 4.36으로 분석되었다.

요약하면, 제주도의 전반적인 만족도에 관한 관광 빅데이터와 스몰데이터를 통해 비교·분석한 결과, 전반적인 만족도는 유의적인 차이가 없음을 확인 할 수 있다. 이를 보다 구체적으로 살펴보면, <표 6>에 제시된 바와 같다.

5) 분석결과 종합

본 연구의 분석결과를 요약하여 설명하면 다음과 같다. 관광 빅데이터 분석(1차)에서 경치(1위, 0.5027), 가족여행(2위, 0.4834), 드라이브(3위, 0.3635), 힐링(4위, 0.3209), 등산(5위, 0.2629)의 순으로 나타났다. 한

편, 스몰데이터 분석(2차)에서는 경치(1위), 가족여행(2위), 데이트(3위), 드라이브(4위), 사진 찍기(5위)의 순으로 나타났다. 이를 통해 제주도 방문목적은 ‘경치’와 ‘가족여행’의 순으로 나타났고, 두 개의 분석결과 순위에서 일치하고 있음을 알 수 있다. 특히, 1구간에서의 순위의 일치 개수가 3개로 나타났다. 제주도 분위기는 1차 분석에서는 ‘재미있고’, ‘이국적이며’, ‘깨끗한’이라고 분석된 반면, 2차 분석에서는 ‘평온한’, ‘깨끗한’, ‘분위기 좋은’으로 나타나 두 개의 분석결과 순위에서 일치하지 않았다. 종합하면, 1구간에서 분위기가 ‘깨끗하다’, ‘편안하다’가 포함되었음을 알 수 있고, 나머지 구간에서의 순위에는 차이가 나타났음을 알 수 있다. 제주도의 인기토픽에서는 ‘해변’을 제외하고 거의 모든 구간 및 구간 내 순위에서 일치하지 않은 것으로 분석되었다. 특히, 음식 관련 키워드가 두드러지게 순

<표 6> 제주도의 전반적인 만족도 분석결과

구 분	관광 빅데이터 분석 1차: N=223개소	스몰데이터 분석 2차: N=225명
만족도	4.33	4.36

위에 올랐다는 점이다. 이러한 결과는 인기토픽은 방문목적이나 분위기와 달리 단시간에 인기의 흐름 혹은 트렌드가 변화할 수 있다는 점을 확인하였다는 점에서 의의가 있다고 판단된다. 제주도의 전반적인 만족도에 관한 관광 빅데이터와 스몰데이터를 통해 비교·분석한 결과, 전반적인 만족도는 유의적인 차이가 없음을 확인 할 수 있다.

## V. 결론 및 시사점

### 1. 연구요약 및 결론

본 연구는 코로나19와 포스트 코로나 기간 제주도의 방문 목적을 비롯하여 분위기 등 제주도의 전반적인 관광특성에 관한 관광객들의 행동과 인식을 관광 빅데이터와 스몰데이터를 활용하여 분석하고자 하였다. 구체적인 연구 목적은 첫째, 국내 플랫폼 기업의 데이터 랩의 오픈 소스에서 제시하고 있는 제주도 관광자원의 방문목적과 분위기, 인기토픽, 그리고 만족도 분석결과와 제주방문 관광객을 대상으로 한 설문조사인 스몰데이터 분석결과를 비교·분석하고자 하였다. 둘째, 두 개의 분석결과를 항목별로 정리한 후 이를 시각화 기초분석 결과를 제공하여 해당 지역에서 향후 고도화할 수 있도록 제안하였다. 이에 따라, 국내 대표 플랫폼인 네이버의 데이터 랩(DataLab)의 오픈 소스와 제주도를 방문한 관광객을 대상으로 한 설문 조사를 병행하여 활용하였다.

연구결과, 첫째, 제주도 방문목적에 관한 관광 빅데이터(1차)와 스몰데이터(2차) 분석결과, ‘경치’와 ‘가족여행’의 순으로 나타났고, 두 개의 분석결과와의 최상위 순위에서 상호일치 하였으며, 특히, 1구간 내에서의 순위의 일치 개수가 3개로 나타났다. 둘째, 분위기에서 관광 빅데이터 분석에서는 ‘재미있고’, ‘이국적이며’, ‘깨끗한’으로 분석되었으나 스몰데이터 분석에서는 ‘평온한’, ‘깨끗한’, ‘분위기 좋은’으로 나타나 전체 순위에서 완벽히 일치하지 않았으나 ‘깨끗한’의 분위기는 동일 구간(1구간)에 포함되었다. 나머지 구간에서의 전체 순위 및 구간 내 순위에는 차이가 나타났다. 셋째, 인기토픽에서 관광 빅데이터는 ‘오름’, ‘해변’, ‘박물관’으로 분석되었으나 스몰데이터 분석에서는 ‘고기국수’, ‘갈치’, ‘해변’으로 나타나 ‘해변’을 제외하고 전체

순위 및 구간 내 순위에서 일치하지 않은 것으로 분석되었다. 특히, 스몰데이터 분석결과 음식과 관련된 키워드가 두드러지게 순위에 올랐음을 알 수 있었다. 이러한 결과는 인기토픽 항목은 방문목적이나 분위기와 달리 단시간에 인기의 흐름 혹은 수요가 변화할 수 있다는 점을 확인하게 되었다. 넷째, 제주도의 전반적인 만족도를 분석한 결과, 관광 빅데이터는 4.36으로 분석되었고, 스몰데이터는 4.33으로 나타나 두 분석결과 간의 유의적인 차이가 없음을 확인하였다. 종합하면, 두 개의 분석결과에서 제주도 방문목적과 전반적인 만족도는 거의 일치하고 있으나 분위기와 인기토픽은 분석결과 간 전체적인 순위 및 구간 내 순위에서 크고 작은 차이가 나타났음을 확인하였다.

본 연구결과 시사점은 다음과 같다. 첫째, 두 개의 분석결과, 제주방문 목적은 ‘경치’와 ‘가족여행’으로 나타나 경치를 즐기는 가족여행을 위한 방문객이 가장 높게 나타남에 따라 관계기관에서도 무계획적인 난개발을 자제하여 지금과 같은 무공해의 자연환경을 유지하는 것이 무엇보다 필요하다고 판단된다. 또한, 젊은이를 대상으로 데이트나, 드라이브, 사진 촬영 등에 목적을 두고 방문하는 고객을 위한 보다 편리한 교통시설 혹은 프로그램의 개발과 중장년층이 대상이 되는 나들이, 힐링, 등산 등을 위한 재정비 및 다양한 프로그램 개발이 필요하다는 점이다. 둘째, 사회적으로 활발하게 역할을 하게 되는 30~40대가 가장 많이 제주도를 방문하는 부분을 고려하여 이에 적절한 여행상품의 개발, 숙박형태, 식음료 부분의 메뉴개발 그리고 여행의 즐거움에서 빼놓을 수 없는 쇼핑을 즐길 수 있는 다양한 시설에 관심을 두고 접근을 해야 할 것이다. 아울러 점점 더 해외여행으로의 관심을 제주도로 되돌리기 위한 양질의 관광상품 개발과 서비스 개선 노력이 수반되어야 할 것이다. 셋째, 제주도의 인기토픽이 ‘해변’을 제외하고 매우 급격하게 변화하고 있음을 확인하였다. 특히, 음식 혹은 먹방 관련 연관어가 매우 높은 순위를 차지하고 있어 지역특산물을 활용한 특화된 식음료상품의 연구 및 지속적인 메뉴개발과 품질 개선, 그리고 더욱 친절한 서비스 방안과 제공이 필요하다고 본다. 넷째, 관광 빅데이터와 스몰데이터 분석방법을 활용한 관광분야의 초기연구로 시간적 그리고 비용적 측면의 상당한 노력이 요구되지만, 온라인상 정보의 정확성과 신뢰성, 그리고 타당성을 확인할 수 있었으며, 관련



학계에서 이러한 분석방법을 더욱 많이 활용할 수 있는 하나의 계기를 만들었다는 점에서도 학문적 의미가 있었다고 본다.

본 연구의 한계는 다음과 같다. 첫째, 자료수집의 과정 및 연구특성 상 시차가 발생했다는 점이다. 관광 빅데이터는 지난 2021년 9월부터 12월까지 수집되었으나 스몰데이터는 연구설계 과정에서 빅데이터 분석 결과를 바탕으로 스몰데이터 설문지의 문항을 추출하고, 분류하며, 확정하고 난 이후에 본 설문을 진행해야 하는 불가피한 상황이 있었다. 설문은 2022년 9월부터 12월까지 진행함으로써 거의 1년간의 시간적인 편차가 발생하였다는 점이다. 향후에는 동일 연도 또는 분기별 관광 빅데이터와 스몰데이터를 활용한 연구가 진행되어야 할 것으로 판단된다. 둘째, 분석에 활용된 연구대상 및 표본 수에 관한 한계이다. 이는 빅데이터와 스몰데이터의 개념 및 분석상의 특성에 따른 동등한 표본으로 접근할 수 없는 근본적인 한계가 존재하며, 여러 측면에서 연구자의 주관적 생각과 판단에 의존할 수밖에 없는 한계가 있었다.

## 2. 연구제안

다음은 관광 빅데이터와 스몰데이터를 방문목적 등 몇 가지 측면에서 비교분석 한 후 다음과 같이 관광 자원(관광지)의 시각화를 제안하고자 한다. 데이터 시각화는 차트, 그래프 또는 맵과 같은 시각적 요소를 사용해 데이터를 표시하는 프로세스로 이를 통해 복

잡하거나, 용량이 크거나, 숫자 혹은 문자로 표현된 데이터를 처리하기 쉬운 시각적 표현으로 변환하는 작업이다. 데이터의 시각화는 연구자와 의사 결정자가 데이터 간의 관계를 식별하고 숨겨진 패턴이나 추세를 감지하거나 데이터 중심의 의사결정 및 경영전략을 수립하는데 적절하게 활용할 수 있다는 점이다. 이에 따른, 안전지도, 범죄지도, 열지도 등 국내에서도 다양한 형태의 데이터 시각화 사례를 제시하고 있어 소비자 혹은 고객의 정보이해를 높이는 데 중요한 역할을 하고 있다. 관광 관련 분야에는 한국관광공사를 비롯하여 연구원, 일부 지자체에서 인포그래픽, 카드뉴스의 형식으로 특정 데이터를 시각화하는 자료를 보여주고 있다. 본 연구에서는 관광 빅데이터 기초자료를 비교 분석하여 기초지자체 혹은 광역단위의 관광지에 관한 방문목적, 분위기, 인기토픽, 만족도 등을 실시간 혹은 특정 주기로 분석, 시각적으로 표현한 후 이를 관광 정보서비스 형태로 제공할 수 있을 것으로 판단된다. 이를 보다 구체적으로 살펴보면, <그림 5> 나타난 바와 같다.

본 연구를 통해 제안되고 있는 관광지 시각화의 구체적인 제안내용 및 주요편의는 다음과 같다. 첫째, 관광 분위기, 관광 만족지도, 관광 이미지 지도 등을 기존의 관광 안내지도(온라인상의 모바일 형태 포함)를 업데이트하여 관광객들에게 제시함으로써 이러한 시각화 자료를 바탕으로 잠재 관광객들이 해당 지역을 방문하도록 유도하거나 다른 관광객에게 추천 기회를 제공할 수 있을 것이다. 둘째, 지자체별로 매년



참 고: 지도상의 글자크기 및 위치는 연구자가 임의로 표시한 것임(만족도 평균제외).

<그림 5> 제주도 방문목적, 분위기, 인기토픽, 만족도 분석 시각화(예시)



혹은 반년 단위의 주기적 혹은 실시간 빅데이터 분석을 통한 시각화 자료를 계속하여 축적함으로써 향후 잠재 관광수요를 예측할 수 있고, 이를 지자체 단위의 관광정책에 적극적이고 최신 트렌드에 맞게 반영할 수 있다는 점이다. 셋째, 제주도 이외 다른 지방자치단체(광역 시도단위) 및 기초 지자체(시군구) 단위로 연구범위의 확대 및 연구결과의 전파가 가능하다는 점이다. 특히, 국내 인기 관광지 중 빅데이터 자료가 많이 축적된 관광지를 중심으로 확산 가능하며, 인구 감소 및 지방소멸로 위기에 처한 지역의 인구문제 극복방안의 수단으로써 뿐만 아니라 관광정책 혹은 마케팅 및 홍보수단으로 활용이 가능할 것으로 판단된다. 이에 본 연구가 희망하는 최종적인 목적을 달성하는데 그 시발점이 되기를 기대한다.

## 참고 문헌

- Wang Jie (2023). 빅데이터 분석을 통한 코로나 19 이후 스타벅스의 마케팅과 소비자의 인식변화. *관광연구*, 38(1), 1-23.
- 강영옥, 조나혜, 이주윤, 윤지영 & 이혜진 (2019). 경험적 모델과 머신러닝 기법을 활용한 SNS 사용자 분류방법 비교: 플리커 데이터의 관광객 분류 방법. *대한공간정보학회지*, 27(4), 29-37.
- 고진현 (2020). 소셜 빅데이터를 활용한 일본관광에 대한 인식 변화 분석: 일본제품 불매운동을 중심으로. *관광경영연구*, 24(6), 1-20.
- 구본근 (2012). FIVE: Hadoop에서 스몰 데이터를 효율적으로 처리하기 위한 파일 가상 확장 기법. *한국정보기술학회논문지*, 10(10), 69-78.
- 권지혜, 이수진 & 김수정 (2019). 스몰데이터 유형에 따른 시각화 표현 방법 연구: 2000년대 이후 인쇄매체를 중심으로. *커뮤니케이션디자인학연구*, 67, 223-235.
- 박득희, 이계희 & 김태구 (2019). 이원모드 연결망 분석을 활용한 정보원천과 관광객들의 다(多)목적 지 여행행동 간의 관계 조사 제주도 지역을 사례로. *호텔관광연구*, 21(3), 36-48.
- 성보현, 최영석 & 설영훈 (2023). 빅데이터를 활용한 COVID-19 전후 충청북도 경제적 손실추정: 관광객 신용카드 소비액을 중심으로. *MICE관광연구*, 23(2), 147-165.
- 심지수 & 오창승 (2020). Using big data and small data to understand linear parks: focused on the 606 trail, USA and Gyeongchun Line forest, Korea. *한국조경학회지*, 48(5), 28-41.
- 이창재 (2022). 온라인 리뷰에 기반한 직무만족의 결정요인과 추천의도 연구: 빅데이터 분석과 스몰데이터 분석 비교를 중심으로. [박사학위 논문, 경희대학교 대학원]. RISS.
- 임종훈 & 하정우 (2021). 소셜미디어 빅데이터를 분석한 관광특구 인식 조사: 서울시 마포관광특구를 중심으로. *Tourism Research*, 46(3), 263-278.
- 제주관광공사 (2021). *코로나19 전후 제주관광 트렌드 분석*. [https://ijto.or.kr/korean/Bd/view.php?btable=pds&bno=332&pds\\_skin=&p=1&lcate=2](https://ijto.or.kr/korean/Bd/view.php?btable=pds&bno=332&pds_skin=&p=1&lcate=2)
- 조윤희 (2023). 소셜미디어 빅데이터를 활용한 의료관광 인식 및 활성화 방안에 관한 연구: 코로나19 발생 전·후 비교를 중심으로. *호텔리조트연구*, 22(1), 129-144.
- 주정, 윤현 & 윤희정 (2021). 빅데이터를 이용한 유산관광 연관단어 및 소셜네트워크 분석. *관광경영연구*, 25(5), 155-173.
- 한국관광공사 (2021). *‘빅데이터로 똑똑하게’ 지역관광사업 해법시다*. <https://contest.visitkorea.or.kr/viewer/view.kto?id=79316&type=bd>
- El Alaoui, I., & Gahi, Y. (2019). The impact of big data quality on sentiment analysis approaches. *Procedia Computer Science*, 160, 803-810.
- Faraway, J. J., & Augustin, N. H. (2018). When small data beats big data. *Statistics & Probability Letters*, 136, 142-145.
- Garner, B., Thornton, C., Pawluk, A. L., Cortez, R. M., Johnston, W., & Ayala, C. (2022). Utilizing text-mining to explore consumer happiness within tourism destinations. *Journal of Business Research*, 139, 1366-1377.
- Giglio, S., Bertacchini, F., Bilotta, E., & Pantano, P. (2019). Using social media to identify tourism attractiveness in six Italian cities. *Tourism Management*, 72, 306-312.
- Gong, L., Liu, X., Wu, L., & Liu, Y. (2016). Inferring trip purposes and uncovering travel patterns from taxi trajectory data. *Cartography and Geographic Information Science*, 43(2), 103-114.
- Guo, Y., Barnes, S. J., & Jia, Q. (2017). Mining meaning from online ratings and reviews: Tourist satisfaction analysis using latent

- dirichlet allocation. *Tourism Management*, 59, 467-483.
- He, Z., Deng, N., Li, X., & Gu, H. (2021). How to “read” a destination from images? machine learning and network methods for DMOs’ image projection and photo evaluation. *Journal of Travel Research*, 61(3), 597-619.
- i Agustí, D. P. (2021). Mapping gender in tourist behaviour based on instagram. *Journal of Outdoor Recreation and Tourism*, 35, 100381.
- Jardim, S., & Mora, C. (2022). Customer reviews sentiment-based analysis and clustering for market-oriented tourism services and products development or positioning. *Procedia Computer Science*, 196, 199-206.
- Kambatla, K., Kollias, G., Kumar, V., & Grama, A. (2014). Trends in big data analytics. *Journal of Parallel and Distributed Computing*, 74(7), 2561-2573.
- LaValle, S., Lesser, E., Shockley, R., Hopkins, M. S., & Kruschwitz, N. (2013). *Big data, analytics and the path from insights to value*. <https://tarjomefa.com/wp-content/uploads/2017/08/7446-English-TarjomeFa.pdf>
- Lindstrom, M. (2016). *Small Data: The Tiny Clues that Uncover Huge Trends*. St. Martin’ s Press.
- Liu, W., & Ji, R. (2018) Examining the role of online reviews in Chinese online group buying context: the moderating effect of promotional marketing, *Social Sciences*, 7, 141.
- Lyu, J., Khan, A., Bibi, S., Chan, J. H., & Qi, X. (2022). Big data in action: an overview of big data studies in tourism and hospitality literature. *Journal of Hospitality and Tourism Management*, 51, 346-360.
- Ma, S. D., Kirilenko, A. P., & Stepchenkova, S. (2020). Special interest tourism is not so special after all: big data evidence from the 2017 Great American Solar Eclipse. *Tourism Management*, 77, 104021.
- Martinez-Torres, M. D. R., & Toral, S. L. (2019). A machine learning approach for the identification of the deceptive reviews in the hospitality sector using unique attributes and sentiment orientation. *Tourism Management*, 75, 393-403.
- Secchi, P. (2018). On the role of statistics in the era of big data: a call for a debate. *Statistics & Probability Letters*, 136, 10-14.
- Singh. S., & Singh, N. (2012). *Big Data analytics*. <https://ieeexplore.ieee.org/document/6398180>
- Zakir, J., Seymour, T., & Berg, K. (2015). *Big data analytics*. [https://iacis.org/iis/2015/2\\_iis\\_2015\\_81-90.pdf](https://iacis.org/iis/2015/2_iis_2015_81-90.pdf)