



다이내믹 토픽 모델링을 활용한 아프리카 사파리 관광 동향 분석* 탄자니아 영문 온라인 리뷰를 중심으로

Trend assessment on African safari tourism applying dynamic topic modeling: A case of Tanzania online reviews in English

이 재 현** · 정 원 지*** · 김 성 수**** · 정 철*****

Lee, Jaehyun · Chung, Won-Ji · Kim, Sungsoo · Jeong, Chul

요 약: 아프리카 탄자니아는 사파리 관광을 활용하여 사회·경제적 이익을 창출하고 있다. 사파리 관광은 생태자원과 야생동물을 중심으로 이루어지기 때문에 관광 대상에 대한 불확실성이 존재한다. 관광객들은 동향 파악과 관광 대상의 불확실성 감소를 위해 온라인 리뷰를 적극 활용하는 경향이 있다. 이 연구는 탄자니아 사파리 관광 상품에 대한 인식의 주요 토픽과 동향 변화를 알아보는 것을 목적으로 두고 있다. 연구 목적을 달성하기 위해 Safaribookings의 영문 온라인 리뷰 데이터를 스크래핑 하였으며, 연구문제 1과 연구문제 2를 규명하기 위해 시계열적 흐름과 동향 분석에 적합한 다이내믹 토픽 모델링(DTM) 알고리즘을 활용하였다. 분석 결과, 사파리 관광 추천, 야생동물 관광 경험, 이동수단 지불 경험, 호텔 및 항공 예약 경험, 킬리만자로산 관광 경험, 잔지바르 및 셀루스 지역 관광 경험에 해당하는 주요 6개의 토픽을 확인할 수 있었으며 모두 동향 변화가 있는 것으로 나타났다. 이 연구는 다이내믹 토픽 모델링이라는 알고리즘을 활용해 시계열적 한계 극복과 체계적인 프로세스를 적용하였다는 점에서 연구 방법론적 확대와 다이내믹 토픽 모델링이 체계적인 연구방법임을 시사한다. 나아가 분석 결과를 기반으로 사파리 관광 마케터 등과 같은 이해관계자들에게 탄자니아 사파리 관광 마케팅 전략 수립에 있어 도움이 될 수 있을 것이다.

핵심어: 다이내믹 토픽 모델링, 사파리 관광, 탄자니아, 동향 분석, 온라인 리뷰

Received April 24, 2023 Revised July 17, 2023 Accepted July 18, 2023

* 이 연구는 2021년 대한민국 교육부와 한국연구재단의 지원을 받아 수행된 연구임(NRF-2021S1A5B8096026).

This work was supported by the Ministry of Education of the Republic of Korea and the National Research Foundation of Korea(NRF-2021S1A5B8096026).

** 한양대학교 일반대학원 관광학과 박사과정수료. wogus11224@hanyang.ac.kr

Ph.D. Candidate, Department of Tourism, Hanyang University.

*** 한국문화관광연구원 위촉연구원. wonjichung@gmail.com

Research Assistant, Korea Culture and Tourism Institute.

**** 한양대학교 정치외교학과 교수. skim14@hanyang.ac.kr

Professor, Department of Political Science and International Studies, Hanyang University.

***** (교신저자) 한양대학교 관광학부 교수. jeong72@hanyang.ac.kr

(Corresponding author) Professor, Department of Tourism, Hanyang University.

Copyright © The Tourism Sciences Society of Korea 2023. This is an open access article distributed under the terms of the Creative Commons Attribution License(CC BY, <https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction of the work in any medium, provided the original authors and source are properly cited.

ABSTRACT: Tanzania in Africa is generating socio-economic benefits through safari tourism. Given that safari tourism centers on ecological resources and wildlife, tourists may feel less certain about what they will actually see or encounter. Tourists tend to utilize online reviews to reduce uncertainty and gain insights into trends. This study aims to understand the key topics and trend changes in perceptions of Tanzanian safari tourism products. English online review data from Safaribookings were crawled to accomplish the research purpose, and the Dynamic Topic Modeling (DTM) algorithm, which is suitable for temporal trend analysis, was utilized to address the research questions. The analysis revealed the following six major topics, all demonstrating trend changes: safari tourism recommendations, wildlife tourism experiences, experiences in transport payments, hotel and airline booking experiences, Mount Kilimanjaro tourism experiences, and tourism experiences in Zanzibar and Selous. The study results support that the DTM, which is used to overcome temporal limitations and implement systematic processes, is a systematic research method and can help expand the scope of research methodologies. Further, based on the analysis results, this study can aid stakeholders, such as safari marketers, in establishing Tanzania safari tourism marketing strategies.

Key words: Dynamic topic modeling, Safari tourism, Tanzania, Trend analysis, Online review

I. 서 론

오늘날 관광산업은 선진/개발도상국의 경제 성장과 산업 발전에 기여하는 주요 동력으로 작용하고 있다(Ren et al., 2019). 특히 개발도상국은 관광산업을 활용하여 지역 경제 활성화와 고용 창출, 빈곤 감소 등과 같은 다양한 사회·경제적 파급효과를 창출하고 있다(Kyara et al., 2022). 일례로 탄자니아는 경제 성장의 핵심 동력으로 관광산업을 언급한 바 있다(Antonakakis et al., 2016; Curry, 1990). 이는 2019년 기준, 약 85만 명의 근로자가 관광산업에 종사하고 있는 것도 있으나 그보다는 관광산업이 전체 GDP 중 두 번째로 큰 비중(기여율: 17%)을 차지할 만큼 사회·경제적 파급효과가 높기 때문이다(World Bank, 2021).

하지만 최근 아프리카는 코로나19의 여파로 관광산업을 넘어 국가 경제까지 막심한 피해를 입고 있다. 2019년 아프리카는 관광산업을 통해 약 1,690억 달러 규모의 이익을 창출하였으나, 2020년 코로나19의 영향으로 약 870억 달러의 손실(관광부문 전

체 GDP 51% 수준)을 입었다(Statista, 2020). 이에 아프리카는 관광산업의 재건과 발전에 총력을 기울이고 있으며, 최근에는 아프리카 7개국을 중심으로 한 사파리 관광 프로그램 wilderness safaris 등과 같은 국가교류 관광상품이 큰 주목을 받고 있다.

이러한 노력은 우리나라-아프리카 관광산업 진출에 많은 시사점을 제공할 수 있다. 현재 아프리카가 세계 관광시장에서 차지하는 비중은 높지 않으나 아프리카는 (1) 높은 성장 잠재력으로 세계 최대 소비시장으로 부상하고 있고 (2) 현재 아프리카는 대륙을 잇는 연결망뿐 아니라 대외 접근성을 높이기 위해 규제를 완화하고 있으며 (3) 코로나19 이후에도 한국-아프리카 간 지자체 관광 교류를 지속적으로 하고 있기 때문이다(김성수, 2023.04.27.). 즉, 세계 여러 국가와 치열하게 경쟁하는 현실점에서, 세계 최대 소비시장인 아프리카를 선점하기 위해 우리나라는 아프리카에서의 관광활동을 보다 적극적으로 활발히 진행해야 할 필요성이 요구되는 시점이라 볼 수 있다.

아프리카는 사파리 관광과 같은 야생동물 관광이나 생태관광이 유명한 대륙이다. 사파리 관광은 다

른 관광 활동과 달리 관광 대상에 대한 불확실성이 높아(김주락, 2018) 온라인 리뷰가 중요한 요인으로 작용되기도 한다. 이러한 이유는 온라인 리뷰가 개인의 의사결정에 있어 불확실성을 감소시켜줄 수 있기 때문이다(이정, 2011). 이처럼 온라인 리뷰가 정보 제공이나 구매 의사 결정, 동향 파악 등 하나의 정보출처원으로 인식됨에 따라 텍스트 마이닝을 적용하여 온라인 리뷰를 분석하는 연구가 진행되고 있다. 이는 텍스트 마이닝이 새로운 지식이나 패턴을 도출하고 특정 현상 파악에 용이한 분석 기법이 기 때문이다(Ha & Geum, 2020).

이와 같이 정량적 평가와 정성적인 평가가 지속적으로 이루어지면서 데이터 속 내재된 유의미한 정보를 탐구하는 여러 방법이 논의되고 있음에도 불구하고 선행연구들(박건철, 이치형, 2019; O'Connor, 2010)은 시계열적 변화를 고려하지 않은 채 단어 빈도수만을 중심으로 수행되어왔다. 즉, 선행연구들은 (1) 개별 토픽이 전체 분석 기간에 걸쳐 변화 없이 유지된다는 전제를 가지고 있으며 (2) 토픽들이 상호 간에 영향을 주며 생성되고 사라지는 시계열적 변화 과정 파악의 어려움과 (3) 전체 분석 기간 중 빈번하게 등장하는 단어만 핵심 키워드로 도출된다는 명확한 한계점을 지니고 있다는 것이다. 이러한 한계점은 관광 동향 파악에 있어 치명적이다. 특히 관광 상품/서비스는 비가시성의 경험재이기 때문에 다른 관광자의 경험에 근거하여 작성된 리뷰에 의존하는 경향이 크다(김도경, 김인신, 2017). 사파리 관광은 주로 온라인 리뷰를 통해 관련 정보가 생성되고 공유되기 때문에 온라인 리뷰를 활용한 세부적인 접근이 요구되는 바이다.

이상의 논의와 한계를 극복하기 위해 이 연구에서는 사파리 관광 전문 플랫폼으로 알려진 Safaribookings의 관광자 온라인 리뷰를 활용하여 탄자니아 사파리 관광 상품에 대한 인식의 주요 토픽을 탐구하고 어떠한 동향을 보이는지 알아보고자 하였다. 연구 목적을 달성하기 위해 연구문제 1과 연구문제 2를

도출하였다. 동향을 파악하려면 데이터의 특성과 시계열적 흐름을 반영하여야 한다. 이를 위해 이 연구에서는 토픽 모델링 기법을 확장시킨 다이내믹 토픽 모델링(DTM)을 활용하여 시계열적인 흐름을 반영해 기간에 따른 동향 변화를 분석하고자 하였다. 다이내믹 토픽 모델링이라는 알고리즘을 활용해 시계열적 한계 극복과 체계적인 프로세스를 적용하였다는 점에서 기존 연구 방법론적 확대와 선행연구와의 차별적인 학술적 의의를 지닌다. 나아가 분석 결과를 기반으로 탄자니아 정부 관광청과 관광 마케터, 관광 에이전시 등에게 프로그램 제안, 마케팅 전략 수립, 가이드라인 제공 등 다양한 실무적 시사점을 제안할 수 있을 것이다. 이 연구의 주요 연구 문제는 다음과 같다.

연구문제 1: 아프리카 탄자니아 사파리 관광 상품에 대한 인식의 주요 토픽은 무엇인가?

연구문제 2: 아프리카 탄자니아 사파리 관광 상품에 대한 인식은 어떠한 동향을 보이는가?

II. 이론적 고찰

1. 탄자니아와 사파리 관광

사파리는 스와힐리어로 장엄한 여행을 뜻하였으나 오늘날에는 아프리카 초원에서 야생동물을 직접 관찰하는 관광활동으로 정의되고 있다(김주락, 2018). 2021년 6월, 코로나19의 여파로 사파리 여행사의 약 60%가 75% 이상의 예약을 잃었으나, 최근에는 경기 회복과 지속가능한 여행에 대한 관심 증가로 사파리 관광 산업이 점차 회복되고 있는 추이다(Buckley & Chauvenet, 2022; Grand View Research, 2023). 현재 전 세계 사파리 관광 시장 규모는 약 333억 7,000만 달러로 평가되고 있으며 2023년부터 2030년까지 연 평균 5.6%의 성장을 기록할 전망이다(Grand

View Research, 2023). 사파리 관광은 주로 탄자니아, 케냐, 남아프리카, 보츠와나, 남비아 등에서 이루어지며 대부분 전문 가이드와 함께 관광활동을 즐긴다(Buckley & Chauvenet, 2022). 사파리 관광은 연인과 친구를 동반한 유형이 주를 이루며, 세계 각지에서 수 많은 관광객들이 방문하나 특히 미국과 유럽 및 아시아에서 많은 관광객이 방문하는 경향이 있다(Skyquest, 2023). 최근 사파리 관광에서의 지속가능성과 책임 관광에 대한 관심이 증가하면서, 사파리 관광객들의 80%는 지속 가능한 여행을 위해 더 많은 비용을 지불할 의사가 있는 것으로 나타났다(Grand View Research, 2023).

한편 사파리 관광의 대표 국가로 알려진 탄자니아는 2022년 약 125만 명의 관광객이 사파리 관광을 위해 방문하였다(Statista, 2022). 탄자니아가 관광을 통해 벌어들인 수입은 약 25억 6,070만 달러로 나타났으며 (Tanzaniainvest, 2023) 탄자니아에는 현재 800개의 관광 에이전시와 3,750개의 숙박시설(예: 캠핑, 호텔, 등)이 존재한다(Safari Deal, 2023). 하지만 과거 탄자니아는 사회주의 이념으로 인해 탄자니아 정부가 관광산업을 통제하였다. 이는 관광산업의 발전이 자립 정책인 우자마(Ujamaa) 정책에 기여하지 못하고 선진국에 대한 의존도가 높아질 수 있다고 판단하였기 때문이다. 그러던 중 탄자니아 정부는 인접국인 케냐의 관광산업 성장의 영향을 받아 관광 인프라 구축 및 활성화 정책을 실시하였다(Wade et al., 2001). 이러한 영향으로 1951년 1개에 불과했던 국립공원이 현재는 22개에 달하며, 탄자니아 전체 국토 면적의 1/3이 관광과 자연보존을 위해 정부의 보호를 받고 있다(Melubo & Buzinde, 2016; Safari Deal, 2023).

탄자니아는 다양한 천연자원을 보유하고 있어 천연자원을 활용한 다양한 관광 활동이 이루어지고 있다. 탄자니아의 주요 관광 활동으로는 야생동물 관광이 있으며, 야생동물 관광은 자연 서식지에서 현지 야생동물과의 상호작용을 중심으로 하는 관광

활동을 의미한다(Rizzolo, 2023). 야생동물 관광은 관찰이나, 사진 촬영, 사냥 및 낚시 등 다양한 활동을 포괄하는 개념으로(Newsome et al., 2005), 야생 동물 관광의 대표적인 사례로는 사파리 관광이 있다.

사파리 관광이 사회·경제적 파급효과(예: 고용 창출과 경제 성장, 빈곤 감소)를 야기함에 따라 탄자니아에서는 사파리 관광을 활용하여 국내·외 관광객을 유치하는 경향이 있다(Manrai et al., 2020; Wardle et al., 2021). 탄자니아 사파리 관광객들은 주로 유럽, 아시아, 미국에서 방문하며 평균 체류 기간은 7일에서 10일 정도로 알려져 있다(Safari Deal, 2023; Tanzaniainvest, 2023). 여행 스타일과 기간, 프로그램 별 상이하지만 탄자니아 사파리 관광은 일반적으로 1인 당 하루 평균 약 US\$ ~300 수준이며, 럭셔리 패키지일 경우 US\$ 600 이상이 요구된다(Safaribookings, 2022; Safari Deal, 2023). 사파리 관광이 국립공원을 중심으로 이루어지다 보니 5개의 주요 국립공원(세렝게티, 킬리만자로, 타랑기레, 만야라 호수, 아루샤 국립공원)에서 탄자니아 연간 총 수입의 약 95%를 차지하고 있다(Ranke et al., 2023). 이 중에서도 세렝게티 국립공원과 미쿠미 국립공원이 관광객들 사이에서 큰 인기를 끌고 있는데 이는 자연경관이 수려하고 Big 5 출몰 가능성이 높은 국립공원이기 때문이다(Melita & Mendlinger, 2013).

Big 5는 아프리카 사자(panthera leo)와 표범(panthera pardus), 아프리카 코끼리(loxodonta africana), 버팔로(syncerus caffer) 그리고 검은 코뿔소(black rhino)를 말한다. Big 5는 19세기 아프리카에서 동물 사냥을 하던 백인 사냥꾼들이 사냥하기 어렵고 위험한 동물들을 꼽은 데에서 유래되었다(Williams et al., 2000). 사파리 관광객들은 주로 Big 5를 보기 위해 국립공원에 방문하는 경향이 있는데, 이는 사파리 관광객들이 Big 5 동물들을 경외하고 매력 있는 요인으로 인식하고 있기 때문이다(Jakarta Globe, 2019; National Geographic, 2019a).

한편 국내의 관광 연구 분야에서 사파리 관광에 대한 연구가 이루어지고 있다. Shang and Luo (2022)는 사파리 관광 경험을 개선하기 위하여 트립어드바이저 온라인 리뷰를 활용해 사파리 관광의 주요 품질 요인 도출한 바 있다. 또한 Matolo and Salia (2023)는 탄자니아 세렝게티 국립공원을 방문하는 해외 관광객을 대상으로 목적지 충성도에 대한 결정 요인을 탐구하였으며 경쟁력 확보를 위해 공원 입장료나 관광객에게 부과되는 비용을 검토하여 가치를 높일 필요성이 있다고 언급하였다. 한편 국내에서는 다음과 같은 연구가 이루어지고 있다. 김주락 (2018)은 케냐 마사이라 국립 보호구 사파리 관광을 중심으로 산업, 관광 목적지, 관광객 측면에서 현대 관광 산업의 특성을 확인하고, 정책과 활동이 관광 행위에 중요한 영향을 미친다는 것을 밝혔다. 문헌들을 종합해보면, 사파리는 아프리카 초원에서 자동차를 타고 야생동물을 직접 관찰하는 관광활동으로 볼 수 있으며 수려한 자연경관이나 Big 5 출몰 유무에 따라 관광객의 선호도가 달라지는 것을 알 수 있다.

2. 온라인 리뷰의 개념과 특성

온라인 리뷰는 소비자들이 상품이나 서비스에 대한 사용 후기를 온라인 상에서 주관적으로 평가한 것이다(Bickart & Schindler, 2001). 온라인 리뷰는 오프라인 리뷰에 비해 소비자 정보탐색, 의사결정, 기업 이미지 형성 등에 큰 영향력을 행사한다(Martin & Lomax, 2011). 이는 온라인 리뷰가 (1) 시·공간적 제약이 적어 정보 전달이 용이하고(Xie et al., 2011) (2) 언제든지 상품/서비스에 대한 생각이나 의견을 표출할 수 있으며(Guy et al., 2017) (3) 고객의 선호도나 만족 등을 즉각적으로 파악할 수 있기 때문이다.

온라인 리뷰는 네트워크 기술과 사전 정보를 활용하여 구매 결정이 이뤄지는 산업 즉, 경험재 산업

군의 소비자 의사결정에 큰 영향력을 발휘한다(황해정 외, 2016). 이러한 이유는 온라인 리뷰가 기업에서 일방적으로 제공하는 정보가 아닌, 고객 경험에 의해 자발적으로 이루어지는 정보 공유 활동이기 때문이다(Berezina et al., 2016). 이로 인해 고객은 기업에서 제공하는 정보보다 소비자들이 직접 작성한 온라인 리뷰를 더욱 신뢰하는 경향이 있다(김은미, 2021). 이에 기업/마케터들은 온라인 리뷰와 텍스트 마이닝 기법을 활용하여 고객의 인식이나 패턴, 추세 등을 파악한다(Chen et al., 2022).

그러나 온라인 리뷰와 텍스트 마이닝 기법에는 다음과 같은 한계점들을 가지고 있다(Cambria & White, 2014; Liu, 2012; Tsytsaru & Palpanas, 2012). 대표적으로 (1) 신뢰도와 타당도에 대한 불확실성이 있다. 온라인 리뷰는 익명의 사람들 사이에 이루어지는 커뮤니케이션이기 때문에 그 신뢰성을 판단하기에는 한계가 있다. 이러한 문제는 데이터의 품질 문제와도 연관이 있으며, 오류나 불완전한 비정형 데이터는 분석의 정확도를 저하시킬 수 있다. 또한 (2) 언어의 다양성과 의미의 중의성(어휘 / 구문적)에 대한 문제도 안고 있다. 텍스트의 의미는 문맥에 따라 바뀔 수 있다. 특히 문법과 다양한 표현 방식은 문장을 해석하는 데 매우 중요한 요소이다. 하지만 텍스트 마이닝 기법은 표면적인 의미만을 해석하는 데 주로 활용되며, 맥락과 의도까지 이해하는 데에는 한계가 있다. 추가적으로 (3) 표본의 편향성에 대한 한계점도 지니고 있다. 모든 소비자가 온라인 리뷰를 남기지 않는다. 온라인 리뷰를 남기는 소비자는 대체로 극단적인 경험을 한 소비자일 가능성이 높기에 모든 소비자의 견해를 반영하기에는 어려움이 있다. 이러한 영향은 데이터 분석 시 부정확한 결과를 초래하며, 이외에도 정보제공자와의 추가 정보 획득의 어려움과(Duan, et al., 2008) 온라인 소비자들은 매일 방대한 양의 정보에 노출된다는 단점도 가지고 있다(이진희 외, 2011).

그럼에도 불구하고 온라인 리뷰와 텍스트 마이닝

기법을 활용하는 이유는, (1) 대량의 데이터를 처리할 수 있으며 (2) 비정형 데이터에서 인사이트를 추출하고 (3) 실시간으로 데이터를 분석하고 동향을 파악하는 데 적합하기 때문이다(Cambria & White, 2014; Liu, 2012; Tsytsaru & Palpanas, 2012). 특히 관광산업은 경험재 산업군이기에 온라인 리뷰의 역할이 매우 중요하다. 온라인 리뷰를 활용해 관광자 별 특성 파악과 인식, 맞춤형 서비스 제공 등 다양한 인사이트를 얻을 수 있기에 온라인 리뷰의 활용도와 중요성이 증가하고 있다(이현주, 2017). 실제로 관광자들은 여행 의사결정 과정에서 온라인 리뷰를 적극 활용하는 경향이 있다(김도경, 김인신, 2017). 이에 트립어드바이저와 같은 몇몇 온라인 웹사이트들은 온라인 리뷰의 신뢰성을 높이기 위해 리뷰어들에게 개인정보 등과 같은 정보 공개를 요구하기도 한다(Xia & Bechwati, 2008).

이처럼 온라인 리뷰에 대한 관심과 그 중요성이 증대됨에 따라 관광에서도 온라인 리뷰를 활용한 연구가 수행되고 있다. O'Connor (2010)는 영국에 위치한 100개의 호텔을 대상으로 관광객 평점, 리뷰, 호텔의 가격정보를 활용하여 사용자 제작 콘텐츠(UGC)가 호텔의 평판에 긍정적 역할을 한다는 것을 규명하였다. 또한 Chung et al. (2022)은 뮤직 스트리밍 서비스를 이용하는 이용자의 온라인 리뷰를 활용해 고객만족 영향요인과 구성요인을 탐구한 바 있다. 국내 연구에서는, 박수지 외 (2017)는 네이버 연관 검색어 서비스를 활용하여 관광 관련 온라인 리뷰 데이터를 수집하고, 텍스트 마이닝을 활용하여 관광자 수 추이와 유의성 기반 수요예측을 진행한 바 있다. 기존 연구들을 종합해봤을 때 온라인 리뷰는 고객이 온라인을 활용하여 상품이나 서비스에 대한 경험이나 평가로 볼 수 있으며, 온라인 리뷰가 관광자의 관광지 이미지, 목적지 결정 요인 등에 지대한 영향을 미치는 것을 알 수 있다.

3. 다이내믹 토픽 모델링

토픽 모델링은 통계적 분석 방법으로, 핵심 토픽과 그 의미를 파악하기 위해 문서 내 토픽을 활용하는 방법이다(Blei, 2012). 토픽 모델링은 토픽-키워드 비율이나 키워드가 특정 토픽에 포함될 확률 등을 파악할 수 있어 의미론적 분석기법으로도 인식되고 있다(Leydesdorff & Nerghe, 2017). 토픽 모델링에는 다양한 알고리즘이 있으나 주로 확률론적 기반 알고리즘인 잠재 디리클레 할당(latent dirichlet allocation: LDA)를 활용하여 진행되고 있다(김정규, 정철, 2019).

LDA는 특정 토픽에 대해 다양한 확률 조합을 가지고 있다는 전제를 가지고 있어 각 문서 내 토픽의 확률 들을 알아낼 수 있다는 특징이 있다(Blei, 2012). 하지만 LDA는 시계열적 요소들을 고려하지 못해 시간의 흐름에 따라 변화하는 현상을 심층적으로 파악하기에는 어려움이 있다. 이로 인해 LDA는 문서의 토픽을 추론할 때 주로 사용되고 있다(고은지, 최선영, 2021). 만약 토픽의 동향을 파악하기 위해 시간별로 구분하여 LDA 분석을 하게 되면, 각 시간별 토픽이 서로 독립적이게 되면서 시간에 따른 변화를 추적하기 어렵게 된다. 반면 시계열적 흐름을 고려하지 않고 전체 구간에 대해 LDA 분석을 실시한다면 특정 구간에서 어떤 토픽은 사라지게 될 것이다(이선민 외, 2021). 이러한 한계점을 보완하기 위해 시계열적 요소를 고려한 다이내믹 토픽 모델링(dynamic topic modeling: DTM)이 개발되었다(Blei & Lafferty, 2006).

다이내믹 토픽 모델링은 시간의 흐름에 따라 토픽 변화를 추론하는 통계적인 방법으로, 기존 토픽 모델링(LDA)의 확장된 알고리즘이다. 이 알고리즘은 토픽의 시간에 따른 경향성을 가정하고 이에 따라 주요 토픽을 추출함으로써 시간에 따른 토픽 변화를 체계적으로 파악할 수 있게 한다. 구체적으로 다이내믹 토픽 모델링은 (1) 최초 기간의 데이터로

부터 주요 토픽을 식별하고 (2) 이후 기간을 중심으로 토픽 도출 과정에서 직전 기간의 토픽 정보를 반영하여 (3) 토픽의 변화와 흐름을 관리하는 프로세스로 진행된다. 이러한 프로세스로 인해 다이내믹 토픽 모델링은 대규모 데이터에 대한 동향이나 특정 트렌드를 파악하는 데 효과적이다(이한아름 외, 2023; Blei & Lafferty, 2006; Sha et al., 2020).

하지만 기존 선형적 기준에 근거해 토픽을 분류하는 전통적인 방법과 달리 키워드 중심 통계적 확률에 기반하기에 핵심 토픽을 도출하고 해석하는 데 있어 맹점이 존재한다. 특히 뉴스 도메인처럼 시기에 따라 토픽의 변동성이 큰 동향 분석을 하기에는 한계가 있다(연진욱 외, 2022). 그럼에도 불구하고 다이내믹 토픽 모델링은 대규모의 데이터에 대해 표본 추출 없이 전체 자료를 분석할 수 있고, 분석한 특성에 대해 거시적인 관점에서 탐색적으로 조망할 수 있어(이현숙, 이대영, 2021). 국내외 동향 및 트렌드 관련 연구에서 활발히 활용되고 있다.

Sha et al. (2020)은 다이내믹 토픽 모델링을 활용하여 코로나19 발병 전과 후를 이슈를 비교하였다. 코로나19 관련 트위터 온라인 리뷰 데이터를 기반으로, 네트워크 구축과 국가 비상사태 시 정책 변화추이를 분석하였다. Müller-Hansen et al. (2021)은 다이내믹 토픽 모델링 기법과 독일 연방의회 토론 의정서를 활용해 석탄 발전과 광업에 대한 석탄 관련 정책 동향을 분석한 바 있다. 국내에서는 김동훈 외 (2021)는 국내 블록체인 관련 전반적인 동향과 트렌드를 파악하기 위해 다이내믹 토픽 모델링을 활용하였다. 분석 결과를 토대로 정책 수립과 블록체인 관련 기술 개발 계획에 활용될 수 있다는 점에서 실질적인 함의를 시사한 바 있다. 또한 이한아름 외 (2023)는 다이내믹 토픽 모델링을 활용하여 관광경험에 대한 감정의 총체는 긍정적일지라도 관광경험의 세부 내용이 모두 긍정적이지 않다는 것을 확인하였다. 분석 결과를 기반으로 실무적 관점에서 평가 내역을 살펴 만족도를 제고할 필요성이

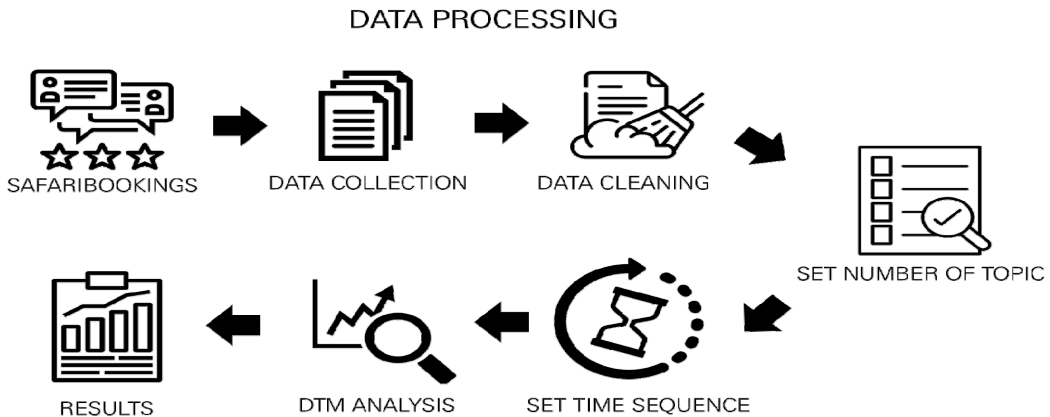
있다고 하였다. 선행연구를 기반으로 종합해보면, 다이내믹 토픽 모델링은 시계열적 특성을 고려하여 문서 속 주요 토픽을 추출하여 그 의미를 파악해주는 기법임을 알 수 있으며 동향이나 트렌드를 분석하는 데 있어 연도별 트렌드를 전체적으로 조망하는 것에 효과적임을 알 수 있다.

III. 연구방법

1. 연구 프레임워크

이 연구는 탄자니아 사파리 관광 상품에 대한 인식의 주요 토픽 탐구와 동향 파악에 목적을 두고 있다. 연구 목적 달성을 위해 사파리 관광 전문 사이트로 알려진 Safaribookings에서 리뷰를 스크래핑하였으며, Python 3.9.7을 활용해 분석을 진행하였다. 데이터 분석에 들어가기 앞서 유의미한 데이터를 추출하기 위해 결측치가 있는 데이터를 제거하고 불필요한 정보를 소거하는 전처리 작업을 진행하였다. 또한 특수문자나 문장부호, 숫자 등과 같이 리뷰 내용에 영향을 미치지 않는 단어들도 제거하였다. 이후 NLP 처리 패키지로 알려진 NLTK(natural language toolkit) 라이브러리를 활용하여 불용어 제거 과정을 진행하였다. 이 과정에서 Tanzania와 Safari 등과 같은 단어를 추가로 제거하였다. 이는 탄자니아 사파리 관광에 대한 리뷰를 수집한 것이기에 빈도의 수나 동향 파악에 있어 의미론적 특성이 적을 것이라고 판단하였기 때문이다.

불용어 제거 이후 관광 업체나 에이전시 등은 연구자가 사전을 활용해 고유명사로 지정하였다. 뿐만 아니라 합성명사 오류를 최소화하기 위하여 연구자는 ‘명사+명사’ 형태의 바이그램(bigram) 형식으로 치환하는 과정을 거쳤다. 이후 형태소 분석에서는 일반 명사, 고유명사, 동사, 형용사만을 추출하였으며, 토픽의 일관성과 적합성을 고려하기 위



〈그림 1〉 데이터 분석 과정 흐름도

해 선행연구(Newman et al., 2010; Roder et al., 2015)를 기반으로 토픽의 수를 결정하였다. 추가적으로 분류된 토픽의 리뷰의 내용까지 확인하여 보다 유의미한 데이터를 추출하고자 하였다. 타임 시퀀스(t)는 2016년부터 2022년까지 1년 단위로 설정하여 연도별로 나타나는 주요 동향을 파악하고자 하였다. DTM 분석은 Gensim의 LDA 시퀀스 모델을 활용하였으며 세부 내용은 〈그림 1〉과 같다.

2. 데이터 수집

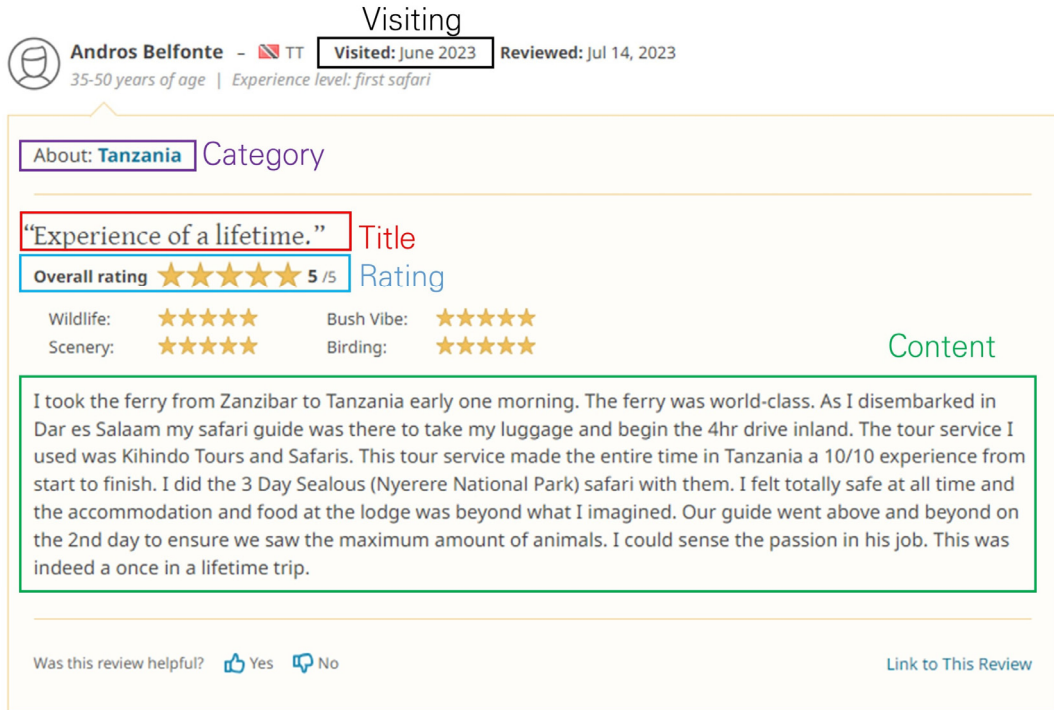
연구 목적을 달성하기 위해 이 연구에서는 SafariBookings(www.safaribookings.com)의 리뷰데이터를 활용하였다(〈그림 2〉 참조). 사파리 관광을 전문으로 하는 플랫폼들이 다수 존재하나, Safaribookings를 선택한 이유는 (1) Safaribookings가 세계 최대 아프리카 사파리 관광정보 공유 사이트로 알려져 있으며, (2) 특정 국가나 지역이 아닌 다양한 사파리 관광 상품을 판매하고 있고, (3) 신뢰성을 위해 리뷰어들에게 거주 국가, 방문일자, 리뷰일자 등의 개인정보를 공개 하도록 되어 있으며 (4) 그 데이터 수 또한 사파리 관광 상품의 수는 약 6,255개, 총 리뷰는 81,089건에 달하기에 사파리 관광 전문성과 적합성을 갖춘 사이트라 판단하였다.

이 연구에서는 데이터 수집을 위해 Octoparse 8을 활용하여 스크래핑 과정을 실시하였다. Octoparse 8은 데이터 크롤링을 전문적으로 하는 웹 사이트이며 이 연구에서 실시한 데이터 수집 방법은 다음과 같다. 먼저 〈그림 2〉에서 보이는 것처럼 (1) 카테고리(category)에서 탄자니아와 관련된 프로그램만을 중심으로 선정하였으며 (2) 평점(rating)과 (3) 제목(title), 그리고 (4) 리뷰 내용(content), 마지막으로 (5) 방문일자(visiting) 위주로 스크래핑 하였다. 방문 일자 는 2016년부터 2022년으로 선정하였으며 수집된 리뷰의 총 수는 4,143건이다.

IV. 분석결과

1. 단어 빈도분석 결과

사파리 관광 키워드를 중심으로 수집된 비정형 데이터를 토대로 연도별 관련된 단어 빈도분석 결과는 〈표 1〉과 같다. 단어 빈도분석 결과, 사파리 관광객들은 2016년부터 2022년 기간 동안 day, great, guide를 가장 많이 언급하였다. 이어서 experience, good, time, serengeti, tour, recommend 등의 키워드 순으로 나타났다. 특히 사파리 관광에 있어 great,



〈그림 2〉 SafariBookings 리뷰 예시

experience, good, recommend, excellent 등과 같은 키워드는 사파리 관광 만족 경험 측면의 단어로 주목해 볼 수 있다. 또한 2020년과 2021년은 코로나19의 직·간접적인 영향으로 인해 사파리 관광 리뷰와 단어 빈도수가 다른 연도에 비해 대폭 감소한 것을 분석을 통해 알 수 있었다(표 2) 참조). 2022년에는 코로나19 방역 규제 완화 등으로 사파리 관광이 다시 상승 추이임을 알 수 있었다.

2. 토픽 모델링 최적화 분석 결과

다이나믹 토픽 모델링 분석에 앞서 토픽의 수를 설정하기 위해서는 혼란도(perplexity)와 응집도(coherence)의 값이 중요하다. 혼란도는 특정 확률 모델이 실제로 관측되는 값을 예측할 때 주로 활용되며 수치가 낮은 값일수록 최적의 결과치로 볼 수 있다. 반면 응집도는 토픽 안에서 의미론적으로 유사한 단어 간

의 유사도를 계산한 값으로, 높은 값일수록 양질의 데이터 분석 결과를 도출할 수 있다. 하지만 혼란도와 응집도는 절대적인 방법은 아니기에 이 연구에서는 도출된 값들을 기반으로 토픽의 의미론적 적합성을 함께 고려하였다. 토픽의 일관성을 설정하기 위해 선행연구(Newman et al., 2010; Roder et al., 2015)에서 제안한 일관성 지수를 기준으로 분석하였다.

토픽 개수에 따른 일관성 지수를 분석한 결과 Topic 5, Topic 6, Topic 10일 때 높은 일관성을 가지는 것으로 분석되었다. 토픽 개수에 따라 도출된 토픽들과 단어들을 비교 분석한 결과 Topic 6의 경우 의미론적으로 적합성이 높은 것으로 나타나 연구자는 토픽 수를 6개로 설정하였다. 다이나믹 토픽 모델링 최적화 분석 세부 결과는 〈표 3〉과 같다.

〈표 1〉 사파리 관광 단어 빈도분석 결과

순위	2016	2017	2018	2019	2020	2021	2022
1	day (344)	great (383)	great (492)	guide (601)	day (182)	guide (170)	guide (496)
2	great (290)	guide (375)	guide (478)	experience (546)	experience (162)	experience (167)	experience (476)
3	experience (270)	day (355)	day (460)	great (541)	guide (142)	great (158)	great (452)
4	guide (258)	experience (331)	good (421)	day (479)	good (137)	day (133)	day (413)
5	time (244)	good (277)	experience (420)	good (459)	time (136)	time (127)	time (305)
6	good (244)	time (266)	time (407)	time (408)	great (132)	good (95)	good (295)
7	serengeti (183)	tour (237)	serengeti (268)	tour (320)	tour (87)	recommend (80)	serengeti (230)
8	driver (164)	company (192)	tour (256)	serengeti (316)	many (79)	serengeti (76)	tour (221)
9	tour (162)	recommend (183)	food (242)	many (266)	driver (78)	food (73)	many (211)
10	recommend (139)	serengeti (182)	excellent (219)	company (245)	recommend (76)	tour (70)	recommend (205)
11	hotel (133)	many (172)	many (218)	driver (239)	company (70)	company (69)	big (178)
12	park (127)	excellent (159)	driver (215)	food (237)	serengeti (69)	amazing (66)	food (174)
13	company (126)	food (158)	park (196)	recommend (221)	park (67)	many (63)	park (165)
14	excellent (118)	driver (131)	company (196)	excellent (199)	food (66)	excellent (56)	company (157)
15	ngorongoro (114)	big (131)	big (184)	see (196)	excellent (64)	wonderful (53)	amazing (154)

주: 괄호()의 숫자는 단어 빈도를 의미함.

〈표 2〉 단어 빈도분석 결과

구분	표제어 추출 후 단어의 수	토큰화 후 단어의 수	불용어 제거 후 단어의 수	품사 태깅 후 도출된 최종 단어의 수
2016	85,333	69,012	34,842	2,931
2017	98,490	79,244	40,538	3,361
2018	137,807	110,042	56,101	4,131
2019	146,693	117,422	59,818	4,085
2020	49,219	39,469	19,957	2,352
2021	43,591	34,770	17,913	2,312
2022	115,941	97,878	46,902	3,614

〈표 3〉 다이나믹 토픽 모델링 최적화 분석 세부 결과

토픽	혼란도	응집도
1	-6.816	.409
2	-6.851	.387
3	-6.858	.407
4	-6.848	.405
5	-6.856	.416
6	-6.869	.431
7	-6.891	.379
8	-6.911	.390
9	-6.961	.408
10	-7.028	.471
11	-7.088	.410

주: 검은색 바탕은 최종 선정된 토픽 수를 의미함.

2. DTM 분석 결과

Topic 1의 리뷰 검토한 결과, 경험, 가이드, 동물, 추천 등에 대한 내용이 주를 이뤘다. 따라서 Topic 1은 사파리 관광 추천에 대한 단어들이 모여 하나의 토픽을 형성하는 것으로 확인되었다. 주요 단어로는 experience, guide, great, animal, recommend 가 있었으며 특히, experience와 guide는 전체기간에서 지속적으로 상위 출현빈도를 갖는 주요 단어로 분석되었다. 그다음으로는 amazing, great, good와 같은 긍정적 평가 순으로 나타났다(〈표 4〉 참조).

2019년과 2020년을 기점으로 guide와 animal, recommend가 상승 축에 비례하여 상승하는 양상은 가이드와 야생동물 관광 경험에 대한 만족으로 추천에 대한 언급이 상승했음을 알 수 있는 부분이다. 특히 사파리 가이드는 다른 관광유형의 가이드와는 달리 야생동물, 문화적인 측면 등을 소개하고 관광객의 안전을 최우선으로 보장하며 전문적인 정보와 해설을 제공해야 한다(Bhuiyan et al., 2021). 이러한 환경으로 인해 사파리 가이드는 가이드 사이의 협력적인 관계를 맺고 있다. 가이드 중 누구라도 Big 5를 발견하면 무전을 통해 다른 가이드에게

연락을 취해 다른 가이드의 관광객도 Big 5를 경험할 수 있도록 한다. 즉, 사파리 관광 가이드들은 사파리 관광객을 위해 장치를 동원하여 관광객의 에로스를 충족시키고자 노력한다(김주락, 2018). 이러한 특징으로 인해 guide, animal, recommend 등의 상승 동향 변화를 기반으로 가이드와 함께하는 사파리 체험 관광 프로그램의 증가가 예상된다.

Topic 2의 리뷰 검토한 결과, 사자, 경험, 야생 등에 대한 내용이 주를 이뤘다. 따라서 Topic 2는 야생동물 관광 경험에 대한 단어들이 모여 하나의 토픽을 형성하는 것으로 확인되었다. 주요 단어로는 see, animal, good, lion, experience, wildlife 가 있었으며, see, animal, lion은 전체기간에서 지속적으로 상위 출현빈도를 갖는 주요 단어로 분석되었다. 그다음으로는 good, amazing, guide 순으로 나타났다(〈표 5〉 참조).

2017년, 2018년, 2020년을 기점으로 lion과 experience, wildlife가 상승 축에 비례하여 상승하는 양상은 관광객들이 야생동물 관광의 경험적 측면을 중요하게 생각하고 있다는 것이 그 이유일 것이다. 특히 Big 5 중 사자에 대한 관심과 경험이 증가 추세임을 알 수 있는데, 이는 탄자니아가 동

〈표 4〉 사파리 관광 추천에 대한 상위 10개 토픽 출현 빈도(Topic 1)

구분	2016	2017	2018	2019	2020	2021	2022
Topic 1	experience (.029)	experience (.029)	experience (.027)	experience (.027)	guide (.026)	guide (.025)	guide (.025)
	guide (.027)	guide (.026)	guide (.026)	guide (.026)	experience (.024)	experience (.023)	experience (.024)
	amazing (.019)	amazing (.018)	great (.018)	great (.022)	great (.019)	great (.016)	good (.015)
	great (.018)	great (.017)	good (.015)	good (.015)	good (.015)	good (.015)	great (.014)
	good (.016)	good (.015)	make (.015)	make (.015)	make (.014)	make (.014)	make (.014)
	make (.015)	make (.015)	amazing (.014)	animal (.013)	animal (.013)	animal (.013)	animal (.013)
	animal (.013)	animal (.013)	animal (.013)	recommend (.012)	recommend (.011)	recommend (.011)	recommend (.011)
	recommend (.012)	recommend (.012)	recommend (.011)	amazing (.011)	amazing (.010)	see (.010)	see (.010)
	see (.011)	see (.011)	see (.011)	see (.011)	see (.010)	go (.010)	go (.010)
	go (.010)	go (.010)	go (.010)	go (.010)	go (.010)	amazing (.009)	amazing (.009)

주: 색칠된 부분은 동향 변화를 가진 주요 토픽을 의미함.

〈표 5〉 야생동물 관광 경험에 대한 상위 10개 토픽 출현 빈도(Topic 2)

구분	2016	2017	2018	2019	2020	2021	2022
Topic 2	see (.030)	see (.031)	see (.032)	see (.033)	see (.034)	see (.036)	see (.038)
	animal (.024)	animal (.024)	animal (.025)	animal (.026)	animal (.026)	animal (.026)	animal (.026)
	good (.013)	good (.013)	lion (.014)	lion (.014)	lion (.014)	lion (.014)	lion (.014)
	lion (.013)	lion (.013)	good (.013)	good (.013)	good (.013)	good (.013)	good (.013)
	amazing (.012)	amazing (.013)	guide (.011)	guide (.011)	great (.011)	experience (.011)	experience (.012)
	guide (.012)	guide (.011)	great (.011)	great (.011)	guide (.011)	great (.011)	great (.010)
	great (.011)	experience (.011)	amazing (.011)	experience (.010)	experience (.011)	guide (.010)	guide (.010)
	experience (.011)	great (.011)	experience (.011)	many (.010)	many (.010)	many (.009)	wildlife (.009)
	wildlife (.010)	many (.010)	many (.010)	amazing (.010)	wildlife (.009)	wildlife (.009)	amazing (.009)
	many (.010)	wildlife (.010)	wildlife (.010)	wildlife (.009)	amazing (.009)	amazing (.009)	many (.009)

주: 색칠된 부분은 동향 변화를 가진 주요 토픽을 의미함.

아프리카에서 사자가 가장 많이 서식하는 지역이기 때문이다(Altezza Travel, 2023). 이러한 특징으로 인해 세렝게티 인근 지역에서는 사자의 대규모 이동을 손쉽게 관찰할 수 있고, lion, experience, wildlife, see, animal 등의 상승 동향 변화를 기반으로 사자를 중심으로 한 사파리 관광 프로그램의 증가가 예상된다.

Topic 3의 리뷰 검토한 결과, 지불, 자동차, 돈 등에 대한 내용이 주를 이뤘다. 따라서 Topic 3은 이동수단 지불 경험에 대한 단어들이 모여 하나의 토픽을 형성하는 것으로 확인되었으며, 주요 단어로는 pay, car, say, take, money가 있었다. 특히, get, pay는 전체기간에서 지속적으로 상위 출현빈도를 갖는 주요 단어로 분석되었으며 그다음으로는 go, car, drive순으로 나타났다(〈표 6〉 참조).

2017년을 기점으로 pay와 say, drive가 하락 추세에 비례하여 하락하는 양상은 지불경험에 대한 이슈가 감소하고 있는 추세임을 알 수 있다. 사파리 관광은 고가의 여행으로 알려져 있어 과거 사파리 관

광 가이드가 비용을 추가로 청구하는 등 가격 투명성에 대한 이슈가 지속적으로 다뤄지고 있었다(Wildwings Safaris, 2015). 이에 사파리 여행사와 관련 이해관계자들은 사파리 가격 투명성을 확보하고자 하고 노력하고 있다. 일례로 2017년 SATSA(Southern Africa Tourism Services Association)와 2021년 TATO(Tanzania Association of Tour Operators)는 사파리 여행사들을 대상으로 가격 투명성을 강조하기 위해 프로그램을 개설한 바 있으며, 여행 가격을 공개하고 관광객에게 명확한 정보를 제공하도록 유도한 바 있다. 이러한 특징으로 인해 pay, say, drive에 대한 하락 동향 변화를 기반으로 이동수단에 대한 지불 이슈는 점진적으로 감소할 것으로 유추할 수 있다.

Topic 4의 리뷰 검토한 결과, 호텔, 기사, 공항, 머뭇, 픽업 등에 대한 내용이 주를 이뤘다. 따라서 Topic 4는 호텔 및 항공 예약 경험에 대한 단어들이 모여 하나의 토픽을 형성하는 것으로 확인되었으며, 주요 단어로는 stay, airport, pick이 있었다. 특히,

〈표 6〉 이동수단 지불 경험 상위 10개 토픽 출현 빈도(Topic 3)

구분	2016	2017	2018	2019	2020	2021	2022
Topic 3	get (.015)	get (.015)	get (.015)	get (.014)	get (.014)	get (.014)	get (.014)
	pay (.015)	go (.014)	go (.014)	go (.013)	go (.013)	go (.014)	go (.014)
	go (.014)	pay (.013)	pay (.011)	pay (.010)	take (.010)	take (.010)	take (.010)
	car (.010)	car (.010)	car (.010)	take (.009)	car (.009)	car (.009)	car (.009)
	drive (.008)	take (.008)	take (.009)	car (.009)	pay (.009)	pay (.009)	hour (.008)
	say (.008)	drive (.008)	drive (.008)	hour (.008)	hour (.008)	hour (.008)	pay (.008)
	take (.008)	say (.008)	hour (.008)	drive (.008)	drive (.007)	read (.007)	read (.007)
	driver (.008)	hour (.008)	say (.008)	say (.007)	read (.007)	drive (.007)	drive (.007)
	money (.008)	driver (.008)	driver (.007)	driver (.007)	say (.007)	driver (.007)	driver (.006)
	hour (.008)	money (.007)	company (.007)	company (.007)	driver (.007)	say (.007)	say (.007)

주: 색칠된 부분은 동향 변화를 가진 주요 토픽을 의미함.

good과 hotel은 전체기간에서 지속적으로 상위 출현빈도를 갖는 주요 단어로 분석되었고 그다음으로는 nice, driver 순으로 나타났다(〈표 7〉 참조).

2017년과 2019년을 기점으로 airport와 book이 상승 축에 비례하여 상승하는 양상은 호텔 및 항공 예약 경험에 대한 언급을 더 많이 했음을 알 수 있다. 아프리카를 중심으로 발병한 에볼라 바이러스(2014~2016)와 코로나19의 영향으로 아프리카 관광산업이 주춤하였으나, 엔데믹을 맞이하면서 수요의 증가로 볼 수 있다. 이러한 특징으로 인해 airport와 book에 대한 상승 동향 변화는 잔지바르 내 호텔 및 항공 예약 서비스가 보다 활성화 될 것으로 유추할 수 있다.

Topic 5의 리뷰 검토한 결과, 킬리만자로, 트레킹, 하이킹, 가이드 등에 대한 내용이 주를 이뤘다. 따라서 Topic 5는 킬리만자로산 관광 경험에 대한 단어들만 모여 하나의 토픽을 형성하는 것으로 확인되

었으며, 주요 단어로는 kilimanjaro, climb, mountain이 있었다. 특히, kilimanjaro와 climb는 전체기간에서 지속적으로 상위 출현빈도를 갖는 주요 단어로 분석되었으며 그다음으로는 mountain, team, guide 순으로 나타났다(〈표 8〉 참조).

2020년을 기점으로 hike와 trek이 하락 축에 비례하여 하락하는 양상은 킬리만자로산에서 이루어지는 관광 경험 중 하이킹과 트레킹에 대한 수요가 낮아지는 것으로 볼 수 있다. 이는 킬리만자로가 아프리카에서 가장 높은 산으로 알려지고 다른 기술이나 장비에 의존하지 않아도 등반할 수 있다는 장점(National Geographic, 2019b)으로 인해 나타나는 현상으로 보인다. 나아가 climb가 상위 출현빈도임을 고려할 때 킬리만자로 등반 관련 프로그램은 꾸준한 수요가 있을 것으로 볼 수 있다.

Topic 6의 리뷰 검토한 결과, 잔지바르와 셀루스 지역, 해변, 휴식 등에 대한 내용이 주를 이뤘다.

〈표 7〉 호텔 및 항공 예약 경험 상위 10개 토픽 출현 빈도(Topic 4)

구분	2016	2017	2018	2019	2020	2021	2022
Topic 4	good (.041)	good (.041)	good (.042)	good (.043)	good (.045)	good (.047)	good (.048)
	hotel (.032)	hotel (.034)	hotel (.034)	hotel (.033)	hotel (.032)	hotel (.031)	hotel (.030)
	nice (.024)	nice (.024)	nice (.023)	nice (.023)	nice (.024)	nice (.023)	nice (.024)
	driver (.022)	driver (.022)	driver (.022)	driver (.022)	driver (.022)	driver (.023)	driver (.023)
	day (.017)	day (.018)	day (.018)	day (.019)	day (.019)	day (.019)	day (.020)
	stay (.016)	zanzibar (.016)	zanzibar (.017)	zanzibar (.016)	zanzibar (.016)	zanzibar (.015)	airport (.015)
	zanzibar (.016)	stay (.016)	airport (.016)	airport (.015)	airport (.015)	airport (.015)	book (.015)
	book (.015)	airport (.015)	stay (.015)	book (.015)	book (.015)	book (.015)	zanzibar (.015)
	airport (.015)	book (.015)	book (.015)	stay (.015)	stay (.014)	stay (.014)	stay (.014)
	pick (.014)	pick (.014)	pick (.014)	pick (.014)	pick (.014)	pick (.014)	pick (.013)

주: 색칠된 부분은 동향 변화를 가진 주요 토픽을 의미함.

〈표 8〉 킬리만자로산 관광 경험 상위 10개 토픽 출현 빈도(Topic 5)

구분	2016	2017	2018	2019	2020	2021	2022
Topic 5	kilimanjaro (.085)	kilimanjaro (.084)	kilimanjaro (.082)	kilimanjaro (.076)	kilimanjaro (.072)	kilimanjaro (.067)	kilimanjaro (.063)
	climb (.050)	climb (.050)	climb (.050)	climb (.050)	climb (.049)	climb (.047)	climb (.045)
	mountain (.038)	mountain (.037)	mountain (.037)	mountain (.037)	mountain (.035)	mountain (.033)	mountain (.032)
	team (.033)	team (.033)	team (.033)	team (.032)	team (.031)	team (.031)	team (.031)
	guide (.028)	guide (.028)	guide (.028)	guide (.028)	guide (.029)	guide (.029)	guide (.029)
	summit (.023)	summit (.024)	summit (.024)	summit (.025)	summit (.025)	summit (.026)	summit (.027)
	hike (.023)	hike (.022)	hike (.021)	hike (.021)	company (.021)	company (.022)	company (.022)
	trek (.020)	trek (.020)	company (.020)	company (.020)	hike (.021)	hike (.020)	make (.020)
	company (.019)	company (.019)	trek (.020)	trek (.019)	trek (.019)	make (.019)	hike (.019)
	crew (.018)	crew (.018)	make (.018)	make (.018)	make (.019)	trek (.019)	trek (.019)

주: 색칠된 부분은 동향 변화를 가진 주요 토픽을 의미함.

따라서 Topic 6는 잔지바르 및 셀루스 지역 관광 경험에 대한 단어들인 모여 하나의 토픽을 형성하는 것으로 확인되었으며, 주요 단어로는 zanzibar, selous, visit, beach가 있었다. 특히, zanzibar와 selous는 전체기간에서 지속적으로 상위 출현빈도를 갖는 주요 단어로 분석되었으며 그다음으로는 visit, local, beach 순으로 나타났다(〈표 9〉 참조).

2021년을 기점으로 selous, visit, beach가 상승 추세에 비례하여 상승하는 양상을 고려할 때 관광객들은 잔지바르와 셀루스 지역을 중심으로 다양한 프로그램에 참여하거나 방문하는 것을 유추할 수 있다. 특히 셀루스는 아프리카 내 대규모 야생 동물 보호구역과 같은 생태 관광이 유명한 지역이다. 코로나19의 영향으로 자연 기반 관광에 대한 관심이 증가하면서(Venter et al., 2020) 다양한 자연경관과 생태계와 관련된 키워드가 함께 상승한 것으로 유추할 수 있다.

V. 논의 및 시사점

오늘날 관광산업은 경제 성장과 지역 산업 발전에 핵심 역할을 하고 있다. 특히 개발도상국은 인적 의존도가 높은 관광산업을 활용해 다양한 사회·경제적 파급효과를 얻고 있다. 이 연구는 아프리카 탄자니아 사파리 관광 상품에 대한 인식의 주요 토픽 탐구와 사파리 관광 상품에 대한 인식이 어떠한 동향을 보이는지 규명하는 데 있다. 연구 목적을 달성하기 위해 아프리카 사파리를 대표하는 탄자니아를 중심으로 2016년부터 2022년까지 Safaribookings에 게시된 총 4,413건의 영문 온라인 리뷰를 활용하였다. 이 연구의 주요 결과는 다음과 같다.

첫째, 이 연구는 아프리카 탄자니아 사파리 관광 동향을 정량적으로 분석하기 위해 2016년부터 2022년까지 Safaribookings에서 수집한 리뷰를 활용하

〈표 9〉 잔지바르 및 셀루스 지역 관광 경험 상위 10개 토픽 출현 빈도(Topic 6)

구분	2016	2017	2018	2019	2020	2021	2022
Topic 6	zanzibar (.090)	zanzibar (.093)	zanzibar (.095)	zanzibar (.097)	zanzibar (.101)	zanzibar (.107)	zanzibar (.113)
	selous (.084)	selous (.080)	visit (.082)	visit (.082)	visit (.081)	selous (.081)	selous (.082)
	visit (.078)	visit (.080)	selous (.077)	selous (.079)	selous (.080)	visit (.080)	visit (.080)
	local (.058)	local (.059)	local (.060)	local (.058)	local (.055)	beach (.053)	beach (.053)
	beach (.055)	beach (.055)	beach (.054)	beach (.053)	beach (.053)	local (.052)	local (.051)
	explorer (.038)	explorer (.040)	explorer (.041)	explorer (.040)	explorer (.039)	explorer (.038)	explorer (.037)
	walk (.028)	walk (.028)	walk (.027)	walk (.027)	walk (.027)	walk (.027)	walk (.027)
	village (.025)	village (.025)	village (.025)	village (.025)	village (.025)	village (.024)	place (.024)
	place (.023)	place (.023)	place (.023)	place (.023)	place (.023)	place (.024)	village (.024)
	island (.020)	relax (.020)	relax (.020)	relax (.020)	relax (.019)	island (.019)	island (.019)

주: 색칠된 바탕은 동향 변화를 가진 주요 토픽을 의미함.

여 아프리카 사파리 관광과 관련된 키워드의 성격을 파악하고자 하였다. 그 결과 상위 15개의 주요 키워드를 도출하였고 이를 통해 어떠한 키워드가 사파리 관광 온라인 리뷰에서 공유되고 있는지 파악할 수 있었다.

둘째, 모델 최적화 분석과 다이내믹 토픽 모델링을 통해 아프리카 탄자니아 사파리 관광 상품에 대한 인식의 주요 토픽은 총 6개의 토픽으로 도출할 수 있었다. 즉, 관광객들이 사파리 관광 활동을 하면서 어떠한 부분을 중심으로 생각하는지, 또 어떤 관광정보가 공유되고 있는지를 고찰할 수 있었으며 아프리카 탄자니아 사파리 관광 상품에 대한 인식의 주요 토픽은 무엇인가에 대한 연구문제 1을 규명할 수 있었다.

셋째, 사파리 관광 추천, 야생동물 관광 경험, 이동수단 지불 경험, 호텔 및 항공 예약 경험, 길리만

자로산 관광 경험, 잔지바르 및 셀루스 지역 관광 경험에 해당하는 토픽 6개의 동향을 확인하였으며 모두 변화가 있는 것으로 나타났다. 이 연구를 통하여 토픽 내 세부 이슈를 시계열적으로 도출할 수 있었으며 토픽과 시간에 따라 변화하는 동향을 고찰할 수 있었다. 이에 아프리카 탄자니아 사파리 관광 상품에 대한 인식은 어떠한 동향을 보이는가에 대한 연구문제 2를 파악할 수 있었다.

이 연구의 이론적 시사점은 다음과 같다. 먼저, 이 연구에서는 내용분석의 한계점으로 지적되어 온 부호화 과정에서의 신뢰도, 타당도 및 다량의 데이터 사용 등 한계점을 보완하여 세부 내용을 파악하였다. 소비자들이 직접 작성한 다량의 온라인 리뷰를 기반으로 텍스트 마이닝 기법을 활용해 연구의 타당성을 확보하여 일반화하고자 하였다는 점에 기존 내용 분석이 지닌 한계점을 극복하고 대체 보완

론적 기법임을 이 연구를 통해 확인할 수 있었다. 또한, 다이나믹 토픽 모델링을 활용해 기존 관광 연구 분야에서 활발히 사용되었던 LDA 분석 기법의 한계를 극복했다는 점이다. 기존 연구(송혜민, 김홍범, 2022)는 개별 토픽이 전체 분석 기간에 걸쳐 변화 없이 유지되는 시계열적 한계점과 빈도수를 중심으로 한 분석이 지배적이었기에 정량적이고 체계적인 방식의 연구 방법론의 필요성이 제기되었다. 이에 이 연구에서는 다이나믹 토픽 모델링이라는 알고리즘을 활용해 시계열적 한계 극복과 체계적인 프로세스를 적용하였다는 점에 연구 방법론적 확대와 기존 관광 연구와의 차별성을 가진다.

실무적 시사점은 다음과 같다. 먼저, 사파리 관광 마케팅과 관광 가이드는 사파리 관광객이 게시한 온라인 리뷰에서 어떠한 관광 경험이 생성되고 공유되고 있는지를 파악하여야 할 필요성이 있다. 관광객들이 최근 선호하는 것은 무엇이고 관광 상품/서비스 개발 및 마케팅 전략 수립에 있어 어떠한 방향성을 기반으로 진행해야 할지 유용한 기초 자료로 제공될 수 있을 것으로 본다. 예를 들어 (1) 야생동물 관광 경험에서 관광객들의 만족 증진을 위해 Big 5 중 사자를 중심으로 한 다양한 야생동물 관광 프로그램 전략 수립이나 (2) 호텔 및 항공 예약 경험에서 예약 서비스 시스템 활성화 전략 그리고 (3) 킬리만자로산 관광경험에서 관광객들의 취향과 선호를 반영한 등산 프로그램 개발 등과 같이 관광객이 관심 있어 하는 최신 니즈를 반영한다면 만족과 재방문율을 높일 뿐 아니라 사파리 관광 수요 증가에도 도움이 될 수 있다. 나아가, 탄자니아 정부 관광청과 관광 업체, 관광 에이전시 등과 같은 이해관계자들은 온라인 리뷰에서 관광객들이 이동수단에 대한 경험을 중심으로 공유하고 있는 것을 주목하여야 한다. 다년간 관광 업체와 관광 에이전시에서 제공하는 관광 서비스에 대한 이슈가 논의되면서 근본적인 문제 해결을 위한 전략이 필요한 시점에 이르렀다. 따라서, 관광 인프라 개선이나 서비스 가이드

라인 보급 등의 물리/사회적 환경을 구축한다면 이는 경쟁력 확보로 이어질 것이며 이미지 전환과 서비스에 대한 이슈 등의 문제 해소에 도움이 될 것이다.

이 연구의 한계점은 다음과 같다. 첫째, 사파리 관광에는 다양한 프로그램이 존재한다는 점이다. 이 연구는 탄자니아라는 지역을 한정하여 사파리 관광에 대한 동향을 조망하였기에 의미 구조상 엮이기 어려운 파편화된 프로그램이 많을 수 있다. 따라서, 후속 연구에서는 프로그램 별로 군집을 형성하여 분석을 진행한다면 편향성 문제를 해결하고 결과의 왜곡을 최소화 할 수 있을 것이다.

둘째, 온라인 리뷰를 기반으로 텍스트 마이닝을 활용해 동향 분석을 수행하였다는 점이다. 온라인 리뷰는 소비자들이 남긴 주관적 평가로 최신 동향에 대해 파악하기 적합하나, 신뢰성과 품질의 문제 그리고 주관성을 배제하기에는 한계가 있다. 따라서 추후 연구에서는 텍스트 마이닝 기법 외에도 다른 분석 기법(예: 시계열 분석 등)을 적용하여 분석을 진행한다면 보다 타당한 연구가 이루어질 것이다.

마지막으로, 온라인 리뷰에는 텍스트 리뷰 이외에도 사진이나 동영상과 같은 다양한 비정형 데이터가 존재한다는 점이다. 이 연구는 텍스트라는 비정형 데이터만을 활용하여 동향 분석을 도출하였다는 점에 제한적이다. 따라서 후속 연구에서는 다양한 비정형 데이터를 활용하여 동향을 비교 분석할 필요성이 제기되며 이를 통해 새로운 연구 토픽 발굴이나 연구의 일반화를 할 수 있을 것이다.

Declaration of competing interest

None.

Author's contribution

Lee, Jaehyun: Conceptualization, Methodology, Data Analysis and Writing(Original draft preparation).

Chung, Won-Ji: Methodology and Writing.

Kim, Sungsoo: Conceptualization and Writing (Reviewing and Editing).

Jeong, Chul: Supervision and Validation, Writing (Reviewing and Editing).

References

고은지, 최선영 (2021). 트윗의 타임 시퀀스를 활용한 DTM 분석: 2019 남북미정상회담 이벤트를 중심으로. <한국정보통신학회논문지>, 25(2), 308-313. <https://doi.org/10.6109/jkiice.2021.25.2.308>

김도경, 김인신 (2017). 텍스트 마이닝을 이용한 온라인 리뷰의 호텔 선택속성 분석. <관광학연구>, 41(9), 109-127. <https://doi.org/10.17086/JTS.2017.41.9.109.127>

김동훈, 오찬희, 주영준 (2021). 다이내믹 토픽모델링 및 네트워크 분석 기법을 통한 블록체인 관련 국내 연구 동향 분석. <정보관리학회지>, 38(3), 23-39. <https://doi.org/10.3743/KOSIM.2021.38.3.023>

김성수 (2023. 04. 27.). 팬데믹 후 '관광산업 부활' 준비하는 아프리카. <내일신문>. http://www.naeil.com/news_view/?id_art=458995

김은미 (2021). 온라인 리뷰의 감성분석과 순환신경망을 적용한 국내 인바운드 관광수요 예측 모형. <관광연구저널>, 35(3), 69-79.

김정규, 정철 (2019). 특허 정보를 활용한 한국과 미국의 관광 관련 기술 동향 분석: 토픽 모델링을 중심으로. <관광학연구>, 43(1), 249-267. <https://doi.org/10.17086/JTS.2019.43.1.249.267>

김주락 (2018). 케냐 마사이마라 국립 보호구의 사파리 관광에서 드러나는 현대관광의 특성. <관광학연구>, 42(7), 11-27. <https://doi.org/10.17086/JTS.2018.42.7.11.27>

박건철, 이지형 (2019). 토픽 모델링을 활용한 스마트시티 연구동향 분석. <인터넷정보학회논문지>, 20(3), 119-128. <https://doi.org/10.7472/jksii.2019.20.3.119>

박수지, 신진옥, 송상현, 정철 (2017). 텍스트 마이닝을 통한 관광지 수요예측: 온라인 검색 엔진을 중심으로. <관광학연구>, 41(1), 13-27. <https://doi.org/10.17086/JTS.2017.41.1.13.27>

송혜민, 김홍범 (2022). 토픽모델링을 활용한 코로나19 관련 관광이슈 및 정부 정책 분석: 정부 정책자료와 온라인 뉴스 비교분석을 중심으로. <관광학연구>, 46(3), 161-182.

연진욱, 부현경, 김남규 (2022). 다이내믹 토픽 모델링의 의미적 시각화 방법론. <지능정보연구>, 28(1), 131-154.

이선민, 천세진, 박상언, 이태욱, 김우주 (2021). 동적 토픽 모델링과 감성 분석을 이용한 COVID-19 구간별 비대면 근무 부정요인 검출에 관한 연구. <정보시스템연구>, 30(4), 277-301.

이정 (2011). 정보이론의 엔트로피 관점에서의 바라본 온라인 소비자 리뷰의 소비자 의사결정에 있어 불확실성 감소 효과. <한국전자거래학회지>, 16(4), 241-256.

이진희, 도선재, 황장선 (2011). 제품 관련 온라인 구전 (e-WOM)의 효과: 방향성, 유용성, 상호작용 성향 (IR)의 역할. <한국심리학회지: 소비자·광고>, 12(2), 283-307. <https://doi.org/10.21074/kjcap.2011.12.2.283>

이한아름, 강혜원, 황선영, 배소영 (2023). 시간의 경과를 고려한 기억에 남는 관광경험 탐구: 다이내믹 토픽 모델링(dynamic topic modeling) 분석을 중심으로. <한국관광학회 제93차 인천국제학술대회 논문집>, 93, 375-381.

이현숙, 이대영 (2021). 다이내믹 토픽 모델링에 기초한 2011-2020 학습분석학 연구 동향 분석. <교육정보미디어연구>, 27(3), 777-815. <https://doi.org/10.15833/KAFEIAM.27.3.777>

이현주 (2017). 빅데이터를 활용한 경복궁 방문 경험 분석: Tripadvisor.com 온라인 리뷰의 활용. <관광연구>, 32(2), 297-318. <https://doi.org/10.17086/JTS.2017.32.2.297.318>

- 21719/KJTR.32.2.14
- 황해정, 심해진, 최준호 (2016). 빅데이터 분석을 활용한 사용자 경험 평가 방법론 탐색: 아마존 에코에 대한 온라인 리뷰 분석을 중심으로. <한국콘텐츠학회논문지>, 16(8), 517-528. <https://doi.org/10.5392/JKCA.2016.16.08.517>
- Antonakakis, N., Dragouni, M., Eeckels, B., & Filis, G. (2016). Tourism and economic growth: Does democrat matter?. *Annals of Tourism Research*, 61, 231-267. <https://doi.org/10.1016/j.annals.2016.09.018>
- Berezina, K., Bilgihan, A., Cobanoglu, C., & Okumus, F. (2016). Understanding satisfied and dissatisfied hotel customers: Text mining of online hotel reviews. *Journal of Hospitality Marketing and Management*, 25(1), 1-24. <https://doi.org/10.1080/19368623.2015.983631>
- Bhuiyan, A. H., Darda, A., & Hasan, R. (2021). Tourists perception and satisfaction on safari tourism at Bangabandhu sheikh mujib safari park in Bangladesh. *International Journal of Geoheritage and Parks*, 9(4), 430-440. <https://doi.org/10.1016/j.ijgeop.2021.11.005>
- Bickart, B., & Schindler, R. M. (2001). Internet forums as influential sources of consumer information. *Journal of Interactive Marketing*, 15(3), 31-40. <https://doi.org/10.1002/dir.1014>
- Blei, D. M., & Lafferty, J. D. (2006). Dynamic topic models. *Proceedings of the 23rd international conference on machine learning*, 113-120.
- _____ (2012). Probabilistic topic models. *Communications of the ACM*, 55(4), 77-84.
- Buckley, R., & Chauvenet, A. L. M. (2022). Hidden economic counterflows in post-COVID international wildlife tourism. *Tourism Management*, 93, 104624. <https://doi.org/10.1016/j.tourman.2022.104624>
- Cambria, E., & White, B. (2014). Jumping NLP curves: A review of natural language processing research. *IEEE Computational Intelligence Magazine*, 9(2), 48-57. <https://doi.org/10.1109/MCI.2014.2307227>
- Chen, T., Samaranayake, P., Che, X., Qi, M., & Lan, Y. C. (2022). The impact of online reviews on consumers' purchasing decisions: Evidence from an eye-tracking study. *Front. Psychol*, 13, 865702. <https://doi.org/10.3389/fpsyg.2022.865702>
- Chung J., Lee, J., & Yoon, J. (2022). Understanding music streaming services via text mining of online customer reviews. *Electronic Commerce Research and Applications*, 53, 101145. <https://doi.org/10.1016/j.elerap.2022.101145>
- Curry, S. (1990). Tourism development in Tanzania. *Annals of Tourism Research*, 17(1), 133-149. [https://doi.org/10.1016/0160-7383\(90\)90119-C](https://doi.org/10.1016/0160-7383(90)90119-C)
- Duan, W., Gu, B., & Whinston, A. B. (2008). Do online reviews matter: An empirical investigation of panel data. *Decision Support Systems*, 45(4), 1007-1016. <https://doi.org/10.1016/j.dss.2008.04.001>
- Grand View Research (2023). Safari tourism market size, share & trends analysis report by type (adventure Safari private Safari), by group (friends, families, couples, solos), by booking mode, by region, and segment forecasts, 2023-2030. <https://www.grandviewresearch.com/industry-analysis/safari-tourism-market-report>
- Guy, I., Mejer, A., Nus, A., & Raiber, F. (2017). *Extracting and ranking travel tips from user-generated reviews*. Proceedings of the 26th international conference on world wide web, 987-996. <https://doi.org/10.1145/3038912.3052632>
- Ha, S., & Geum, Y. (2020). Categorizing sub-categories of mobile application services using network analysis: A case of healthcare applications. *The Journal of Society for e-Business Studies*, 25(3), 15-40.

- Jakarta Globe (2019). Spotting africa's big five in Tanzania. <https://jakartaglobe.id/vision/spotting-africas-big-five-in-tanzania>
- Kyara, V. C., Rahman, M. M., & Khanam, R. (2022). Investigating the environmental externalities of tourism development: Evidence from Tanzania. *Heliyon*, *8*, e09617. <https://doi.org/10.1016/j.heliyon.2022.e09617>
- Leydesdorff, L., & Nerghe, A. (2017). Co-word maps and topic modeling: A comparison using small and medium-sized corpora. *Journal of the Association for Information Science and Technology*, *68*(4), 1024-1035. <https://doi.org/10.1002/asi.23740>
- Liu, B. (2012) *Sentiment analysis and opinion mining*. Morgan&Claypool Publications.
- Manrai, L. A., Lascu, D. N., & Manrai, A. K. (2020). A study of safari tourism in sub-saharan africa: An empirical test of tourism ABC (T-ABC) model. *Journal of Business Research*, *119*, 639-651. <https://doi.org/10.1016/j.jbusres.2019.02.066>
- Martin, D., & Lomax, W. (2011). Word of mouse vs word of mouse: The effects of the internet on consumer's pre purchase. *Occasional Paper Series of Kingston Business School*, *45*, 1-16.
- Matolo, R., & Salia, P. (2023). International tourists' satisfaction with service quality at Serengeti national park in tanzania. *Journal of Co-Operative and Business Studies*, *62*(2), 124-131. <https://journals.mocu.ac.tz/index.php/jcbs/article/view/22>
- Melita, A. W., & Mendlinger, S. (2013). The impact of tourism revenue on the local communities' livelihood: A case of Ngorongoro conservation area, Tanzania. *Journal of Service Science and Management*, *6*(1), 117. <https://doi.org/10.2307/20721420>
- Melubo, K., & Buzinde, C. N. (2016). An exploration of tourism related labour conditions: The case of tour guides in Tanzania. *Anatolia*, *27*(4), 505-514. <https://doi.org/10.1080/13032917.2016.1224976>
- Müller-Hansen, F., Callaghan, M. W., Lee, Y. T., Leipprand, A., Flachsland, C., & Minx, J. C. (2021). Who cares about coal? analyzing 70 years of german parliamentary debates on coal with dynamic topic modeling. *Energy Research and Social Science*, *72*, 101869. <https://doi.org/10.1016/j.erss.2020.101869>
- National Geographic (2019a). What are Africa's big five? Meet the continent's most iconic wildlife. <https://www.nationalgeographic.com/animals/article/africa-big-five-safaris-lions>
- _____ (2019b). Climbing to the roof of Africa with national geographic. <https://www.nationalgeographicexpeditions.co.uk/news/climbing-roof-africa-national-geographic/>
- Newman, D., Lau, j. H., Grieser, K., & Baldwin, T. (2010). Automatic evaluation of topic coherence. *Proceedings of the 2010 Annual Conference of the North American Chapter of the ACL*, 100-108. <https://doi.org/10.5555/1857999.1858011>
- Newsome, D., Dowling, R. K., & Moore, S. A. (2005). *Wildlife tourism*. Channel View Publications.
- O'Connor, P. (2010). Managing a hotel's image on TripAdvisor. *Journal of Hospitality Marketing & Management*, *19*(7), 754-772. <https://doi.org/10.1080/19368623.2010.508007>
- Ranke, P. S., Kessy, B. M., Mbise, F. P., Nielsen, M. R., Arukwe, A., & Roskaft, E. (2023). The threat of covid-19 to the conservation of Tanzanian national parks. *Biological Conservation*, *282*(Wildlife tourism), 110037. <https://doi.org/10.1016/j.biocon.2023.110037>
- Ren, T., Can, M., Paramati, S. R., Fang, J., & Wu, W. (2019). The impact of tourism quality on

- economic development and environment: Evidence from mediterranean countries. *Sustainability*, 11(8), 2296. <https://doi.org/10.3390/su11082296>
- Rizzolo, J. B. (2023). Wildlife tourism and consumption. *Journal of Sustainable Tourism*, 31(5), 1181-1194. <https://www.tandfonline.com/doi/full/10.1080/09669582.2021.1957903>
- Roder, M., Both, A., & Hunneburg, A. (2015). Exploring the space of topic coherence measures. *Proceedings of the Eighth ACM International Conference on Web Search and Data Mining*, 399-408. <https://doi.org/10.1145/2684822.2685324>
- Safaribookings (2022). How much does a Tanzania Safari cost?. <https://www.safaribookings.com/tanzania-safari-costs>
- Safari Deal (2023). Tanzania Safari statistics and facts. <https://safariideal.com/tanzania-safari-statistics-and-facts/>
- Sha, H., Hasan, M. A., Mohler, G., & Brantingham, P. J. (2020). Dynamic topic modeling of the COVID-19 twitter narrative among u.s. governors and cabinet executives. *Social and Information Networks*, arXiv:2004.11692. <https://doi.org/10.48550/arXiv.2004.11692>
- Shang, Z., & Luo, J. M. (2022). Topic modeling for wildlife tourism online reviews: Analysis of quality factors. *Current Issues in Tourism*. <https://doi.org/10.1080/13683500.2022.2086107>
- Skyquest (2023). Global Safari tourism market size, share, growth analysis, by type(couples, and friends), by booking mode(direct bookings, and marketplace booking): Industry forecast 2023-2030. <https://www.skyquestt.com/report/safari-touirm-market>
- Statista (2020). Loss in the contribution of travel and tourism to gross domestic product (GDP) in africa due to the coronavirus (covid-19) pandemic in 2020, by scenario. <https://www.statista.com/statistics/1193582/tourism-gdp-loss-from-covid-19-in-africa/>
- _____ (2022). Number of visitors at national parks in Tanzania as of 2019, by park. <https://www.statista.com/statistics/1248942/most-visited-national-parks-in-tanzania/>
- Tanzaniainvest (2023). Tanzania tourism. <https://www.tanzaniainvest.com/tourism>
- Tsytsarau, M., & Palpanas, T. (2012). Survey on mining subjective data on the web. *Data Min Knowl Disc*, 24, 478-514. <https://doi.org/10.1007/s10618-011-0238-6>
- Venter, Z. S., Barton, D. N., Gundersen, V., Figari, H., & Nowell, M. (2020). Urban nature in a time of crisis: recreational use of green space increases during the COVID-19 outbreak in Oslo, Norway. *Environmental Research Letters*, 15(10), 104075. <https://doi.org/10.31235/osf.io/kbdum>
- Wade, D. J., Mwasaga, B. C., & Eagles, P. F. J. (2001). A history and market analysis of tourism in tanzania. *Tourism Management*, 22(1), 93-101. [https://doi.org/10.1016/S0261-5177\(00\)00019-4](https://doi.org/10.1016/S0261-5177(00)00019-4)
- Wardle, C., Buckley, R., Shakeela, A., & Castley, J. G. (2021). Ecotourism's contributions to onservation: Analysing patterns in published studies. *Journal of Ecotourism*, 20(2), 99-129. <https://doi.org/10.1080/14724049.2018.1424173>
- Wildwings Safaris (2015). Why use a tour operator? <https://wild-wings-safaris.com/blog/why-use-a-tour-operator>
- Williams, P. H., Burgess, N. D., & Rahbek, C. (2000). Flagship species, ecological complementarity and conserving the diversity of mammals and birds in sub-saharan Africa. *Animal Conservation*, 3(3), 249-260. <https://doi.org/10.1111/j.1469-1795.2000.tb00110.x>
- World Bank (2021). Transorming tourism: Toward a sustainable, resilient, and inclusive sector.

<https://documents.worldbank.org/en/publication/documents-reports/documentdetail/794611627497650414/transforming-tourism-toward-a-sustainable-resilient-and-inclusive-sector>

- Xia, L., Bechwati, N. N. (2008). Word of mouse: The role of cognitive personalization in on-line consumer reviews. *Journal of Interactive Advertising*, 9(1), 3-13. <https://doi.org/10.1080/15252019.2008.10722143>
- Xie, H., Miao, L., Kuo, P. J., & Lee, B. Y. (2011). Consumers' responses to ambivalent online hotel reviews: The role of perceived source credibility and pre-decisional disposition. *International Journal of Hospitality Management*, 30(1), 178-183. <https://doi.org/10.1016/j.ijhm.2010.04.008>