

한국관광 실태조사 빅 데이터 분석을 통한 관광산업 활성화 방안 연구*

이정미

한양대학교 일반대학원 경영대학
비즈니스인포매틱스
(jungmi5800@gmail.com)

류미나

한양대학교 일반대학원 경영대학
(mina82400@gmail.com)

임규건

한양대학교 경영대학
(ggseer@gmail.com)

본 연구에서는 한국문화관광연구원에서 조사된 “2013년~2015년 외래 관광객 실태조사”의 약 36,000개 데이터에 대한 빅 데이터 분석을 통해 관광산업 활성화 방안을 도출해 보고자 한다. 이를 위해서 외래 관광객들의 ‘전반적 만족도’, ‘재방문 의사’, ‘추천의사’ 변수에 가장 많은 영향을 끼치는 요인을 분석하고 해당 요인들의 각각에 대한 영향력에 대해 파악 하였다. 본 연구에서는 SPSS IBM Modeler 16.0의 의사결정나무(C5.0, CART, CHAID, QUEST), 인공지능망, 로지스틱 회귀분석의 데이터마이닝 기법을 이용하여 종속변수에 가장 큰 영향을 미치는 상위 변수 7개씩을 각각 도출하였고, 추가적으로 각 독립변수들의 영향력을 심도 있게 파악하기 위하여 R프로그래밍을 활용하여 SPSS IBM Modeler 16.0을 통해 도출된 각 독립변수들의 영향력을 파악하였다. 데이터 분석 결과 ‘전반적 만족도’에 가장 영향을 미치는 상위 변수 7개는 관광지매력도, 음식만족도, 숙박만족도, 교통수단만족도, 안내서비스만족도, 방문관광지수, 국가로 나타났으며 가장 큰 영향력을 미친 변수는 음식만족도와 관광지매력도로 분석되었다. ‘재방문 의사’에 가장 영향을 미치는 상위 변수 7개로는 국가, 여행 동기, 활동, 음식만족도, 제일 좋았던 활동, 관광안내서비스만족도, 관광지매력도로 나타났으며 그중 가장 큰 영향력을 미친 변수는 음식만족도와 여행 동기로 분석되었다. 마지막으로 ‘추천의사’에 영향을 미치는 상위 변수 7개로는 국가, 관광지매력도, 방문관광지수, 음식만족도, 활동, 관광안내서비스만족도, 비용으로 나타났으며 가장 큰 영향력을 미친 변수는 국가, 관광지매력도, 음식만족도로 분석되었다. 따라서 세 변수에 공통적으로 영향을 끼치는 요인은 음식만족도, 관광지매력도로 분석되었으며 해당 요인들이 공통적으로 한국여행에 대한 전반적 만족도와 재방문 의사, 추천의사에 미치는 영향이 크다는 것을 확인할 수 있었다. 본 연구는 외래 관광객들의 한국관광에 대한 활성화 방안을 “외래 관광객 실태조사” 빅 데이터 분석을 통해 규명함으로써 한국 관광 데이터 분석의 활용과 관광 정책 수립의 기초자료로 활용될 수 있을 것으로 기대되며 향후 기업 및 국가차원에서 한국 관광발전에 기여할 수 있는 활성화 방안을 마련하는 자료로 사용될 수 있을 것으로 기대한다.

주제어 : 빅데이터, 인공지능망, 의사결정나무, 로지스틱 회귀분석, R 프로그래밍

논문접수일 : 2017년 11월 6일 논문수정일 : 2018년 4월 11일 게재확정일 : 2018년 6월 5일
원고유형 : 일반논문 교신저자 : 임규건

* The early version of this study was presented on the *Proceedings of the Korea Intelligent Information Systems Society (2016)* (Lee et al., 2016).

1. 서론

현대사회에서 개인의 관광활동은 풍요로운 삶의 질 향상뿐만 아니라 건강증진, 자기 발전, 행복 추구, 경제 활동 등의 다양한 역할을 하고 있다. 관광객은 특정 장소에 일정기간 방문하고 문화적 경험과 체험, 그리고 기억을 가지고 돌아오게 된다(Byun, 2012). 최근 한국뿐만 아니라 전세계적으로 개인의 삶의 질을 높이기 위한 활동으로써 관광은 그 중요성이 더욱 부각되고 있다. Kim and Song(2007)은 관광과 삶의 질의 관계에 관해서 관광경험을 했거나 관광경험에 만족하면 삶의 질이 향상된다고 하였다. 이러한 이유와 더불어 운송수단의 발전과 저가항공의 탄생과 유희활동의 증가로 인한 항공 티켓 가격의 하락으로 여행, 관광 산업은 더욱 활성화 되고 있다. 세계 각국들은“굴뚝 없는 산업”이라고 일컬어지는 관광산업의 발전을 위하여 보이지 않는 관광전쟁을 끊임없이 펼치고 있다. UN WTO에 따르면 전 세계의 국제관광객 수는 1950년 2,500만 명에서 2011년 9억 9천만 명으로 증가하였으며 2012년 10억 명을 돌파, 2030년의 경우 총 18억 명에 달할 것으로 전망하고 있다(Kim, 2014).

우리나라에서는 현재 공공데이터 개방 정책과 “정부3.0”을 기반으로 각 공공기관에서 많은 데이터들을 축적 중에 있다. 그중 특히 관광분야에서 많은 데이터들이 축적되고, 관광객들의 이동과 행위 및 선호와 관련된 관광 빅 데이터를 활용한 연구들이 활발하게 이루어질 전망이다. 관광데이터와 관련해서는 한국문화관광연구원에서 한국관광의 실태를 파악하고 한국관광의 발전을 도모하고자 2007년부터 현재까지 지속적으로 외래 관광객 실태조사를 진행하고 있다. 한국에 방문하는 외래 관광객들 중 매달 약 1,000명

을 조사하여 한국방문의 목적, 여행형태, 여행비용, 개인신상정보 등을 파악한 데이터이다. 하지만 현재까지 데이터의 축적은 많이 되어 있으나 이런 관광 데이터를 활용한 학술적 논의들은 호텔이나 외식(Kong and Song, 2013; Do et al., 2013), 온라인 관광정보의 공공성 분석(Lee et al., 2010), 관광 분야에 SNS 빅 데이터 활용방법 모색(Lee and Yoon, 2014) 정도에 그치고 있어 아직까지 빅 데이터 분석을 통한 활용도는 낮은 편으로 빅데이터를 어떻게 의미 있는 정보로 전환할지 고민이 요구된다(Lee and Yoon, 2014).

본 논문에서는 기존의 연구와 달리 데이터마이닝 기법과 R프로그래밍 기법을 활용하여 빅 데이터 분석을 통해 관광의 만족도에 가장 큰 영향을 미치는 요인을 분석하고자 한다. 기존의 연구들이 외식이나 호텔, 공공성 분석, SNS 분석에 한정되어 있었으나 분석의 범위를 넓힐 수 있다는 차별성을 가진다. 이를 위해서 외래 관광객 실태조사 데이터 약 36,000개에 대한 빅 데이터 분석을 시도하고자 한다. 외래 관광객들의 만족도에 가장 큰 영향을 미치는 요인을 찾고 해당 요인을 보완함으로써 관광발전에 응용할 수 있는 기초자료를 마련하고자 하며 빅 데이터 분석을 통해서 외래 관광객들의 추천의도와 재방문의사에서 영향을 주는 요인들을 도출하고자 한다.

2. 관련연구

2.1 관광 데이터 활용 연구

출입국관광통계 자료에 따르면 국내에 방문한 2016년 외래 관광객은 15,898,425명으로 2015년

에 11,562,192명보다 20.2%가 늘어났으며 2010의 8,797,658명에 비해 거의 배로 증가 하였다(관광정보지식시스템 <http://www.tour.go.kr>). 주요 국가별 입국 비율은 2015년 외래 관광객 실태조사에 따르면 중국인이 47.3%로 제일 많았고 다음으로 일본이 15.6%를 차지하고 그다음 미국, 홍콩, 대만, 태국, 말레이시아, 중동, 싱가포르, 러시아순인 것으로 나타났다.

Parrocoet et al.(2012)는 시칠리아를 방문하는 관광객들의 여행일정 데이터를 사용하여 여행객들이 방문하는 주요 관광지과 관광행태를 규명하고, 이를 통해 현지에 적합한 관광지의 개발전략 및 마케팅 관점의 시사점을 제안하였다. 또한 Kim et al.(2011)은 2009년 외래 관광객 실태조사 데이터를 사용하여 중국인 관광객을 대상으로 그들의 여행행태와 영향요인에 대해 연구를 하였다. 연구결과, 중국인 관광객의 공간 행동은 서울·경기·인천지역의 수도권 방문이 집중적으로 이루어지는 것으로 나타났다. Whang(2015)는 문화체육관광부에서 조사한 2007~2013년 외래 관광객 실태조사를 통해 한국에 방문하는 중국인 관광객의 실태를 파악하고 목적에 따른 수요를 이해하고 관광객의 만족도를 높이기 위하여 “쇼핑관광 활성화”, “상품의 차별화”, “자유여행객 유치”등을 개선해야한다는 방안을 제시하였다.

2.2 빅 데이터 분석

현재 빅 데이터는 사회 전반적으로 주목받고 있으며 정부3.0을 시작으로 축적된 방대한 데이터에 누구든 쉽게 접근할 수 있다. 현재 각 분야에서는 이런 데이터를 이용하여 다양한 분야에서 유의미한 정보를 추출하여 활용하고자 하는 움직임이 활발하다. 이미 많은 공공기업과 지자

체에서는 관광분야에서의 빅 데이터를 활용하고 있다. 이는 관광분야의 특성상 데이터 인프라가 자연스럽게 구축되어 있는 것에 이유를 찾아 볼 수 있다. 이동통신 수단, 신용카드 자료, SNS의 보편화 등에 따라 관광에 관한 정보를 쉽게 구할 수 있게 되었다. 미국의 저장장치업체인 EMC의 발표에 따르면, 2012년을 기준으로 앞으로 2020년까지 현 데이터량의 15배까지 데이터가 증가할 것이라고 발표하였지만 전체 데이터 중 불과 1% 미만만이 분석될 것이라고 예측하고 있다.

빅 데이터는 다양한 형태의 대량 데이터가 실시간 생성되는 대용량(High Volume), 고속도(High Velocity), 고다양성(High Variety)의 특징을 지니고 있다. 빅 데이터는 기업의 프로세스와 조직 등 다양한 문제의 해결책을 제공해 줄 뿐 아니라, 산업 전반과 사회 자체를 변화시킬 수 있는 새로운 방향을 제시하고 있다(Kim, 2013). 이를 활용하여 사용자 로그 및 행태분석, SNS 네트워크 분석, 특정 제품이나 콘텐츠의 추천, 지능형 의사결정 등 다양한 분석을 할 수 있다(Park and Lim, 2015). 최근 들어 관광산업 분야에서도 빅 데이터의 중요성을 인식하고 활용방안과 정부정책 모색 등에도 많은 관심을 보이고 있다.

이러한 중요성을 인식하고 본 연구에서는 빅 데이터를 통해 현재 관광산업의 실태를 규명하고 빅 데이터에 특화된 데이터마이닝 기법을 이용하여 좀 더 수치화된 분석결과를 도출하고자 한다. 본 연구에서는 분류와 예측을 하기 위하여 데이터마이닝 기법 중 의사결정나무, 인공신경망, 로지스틱 회귀분석을 활용하였다. 의사결정나무는 데이터마이닝의 한 기법으로써 보통 대용량의 데이터로부터 그 안에 존재하는 관계와 패턴, 규칙 등을 찾아내 유용한 지식을 추출하는

방법이다. 분석과정이 나무구조로 표현되기 때문에 연구자가 분석과정을 쉽게 이해하고 설명할 수 있는 장점이 있다(Choi et al., 1998). 인공신경망은 생물학의 신경망을 모델링한 알고리즘이다. 인공신경망은 노드들의 그룹으로 연결되어 있으며 이들은 뇌의 방대한 뉴런의 네트워크와 유사하다.

로지스틱 회귀분석은 독립 변수의 선형 결합을 이용하여 하나의 사건이 발생할 가능성을 예측하는 통계 기법이다. 로지스틱 회귀분석의 목적은 일반적인 회귀 분석의 목표와 동일하게 종속 변수와 독립 변수간의 관계를 구체적인 함수로 나타내어 예측모델로 사용된다. 로지스틱 회귀분석은 선형 회귀 분석과는 달리 종속 변수가 범주형인 데이터이며, 입력 데이터에 대해서 분석 결과를 분류하여 제시해 준다.

2.3 관광만족과 재방문에 관한 연구

기존 연구에 따르면 관광만족이란 Stankey (1972)는 ‘방문객들의 동기와 목적의 달성 정도’로 정의하였으며, Hunt(1977)는 ‘최소한 기대되었던 것보다 소비경험이 좋았다고 평가하는 것’이라고 정의하였다. 또한 Pizam(1978)은 ‘관광객의 사전기대와 실제경험 간의 상호작용 결과로 인식하여 여행경험에 대한 특정 영역 관광지 속성에 대한 관광객태도의 조합’이라고 정의하였다. Oh(2012)은 관광만족은 관광객을 관광 상품의 소비자로서 인식하고 상품 소비자의 만족과 불만족에 관한 이론을 관광현상에 적용할 수 있다고 하였다.

또한 Ko(2006)의 연구에 따르면 전반적인 관광객 만족과 함께 재방문은 관광경험을 한 후 심리적 변수라고 하였고 Park(2010)는 관광 후 만

족이 관광전 기대치보다 높으면 관광만족이라고 할 수 있으며 이는 고객창출과 충성고객으로 이어진다고 하였다. Lee and Kim(2014)는 총체적인 관광만족이 관광경험에 대한 전반적 평가라고 할 때 재방문 의도는 관광지경험에 대한 평가라 할 수 있다고 하였다. 또한 재방문 의도는 고객 만족의 정도가 높아질수록 재방문과 구전효과가 커지고 고객의 충성도가 높아지는 것을 의미한다고 하였다(Lee, 2008).

외국인 관광객이 우리나라에 방문하여 한국 고유의 관광경험을 통해 관광만족을 하고 재방문을 하게 되는 것은 우리나라 관광산업에 있어서 매우 중요한 요인이지만 외국인 관광객을 대상으로 관광경험을 통하여 관광만족과 재방문에 관한 연구는 미비한 현실이다(Lee and Kim, 2014; Choi and Park, 2017). 방한 중국인관광객에 대한 전반적 만족도 및 재방문의사는 전반적으로 높게 나타나며, 연령대, 교육수준, 소득, 정보원천 및 직업별로 유의한 차이를 보이고 있다(Kim et al., 2007). 특히 연령에 따라 선택행동 및 관광지선택 기준이 다르게 나타난다고 분석하였다(Park and Kang, 2009). 관광빅데이터 정보 분석은 최근 들어 시도되고 있다. SNS 빅데이터를 활용한 키워드 빅데이터 (Oh et al., 2015; Lee et al., 2014), 웹 검색 트랙픽 정보 분석 (Choi and Park, 2017), 제주도 관광인식에 대한 분석 (Lee et al., 2018) 등 다양한 분야에서 시도되고 있다.

본 연구에서는 외국인 관광객이 실제 설문 응답한 데이터를 바탕으로 관광만족과 재방문의사 그리고 추가적으로 추천의사에 가장 영향을 끼치는 요인에 대해 빅데이터 분석을 통해 파악하고자 한다.

3. 연구방법

3.1 연구 절차

본 연구의 연구 프로세스는 <Figure 1>과 같다. 첫째, 2013년부터 2015년까지의 관광지식정보시스템에 저장된 한국문화관광연구원에서 진행한 외래 관광객 실태조사 데이터를 통합하였다. 둘째, 통합된 데이터를 데이터 정제과정을 통하여 연구목적과 부합하지 않는 불필요한 변수들을 제거하고 분석의 정확성을 높이고자 일부 변수를 조정하여 분석을 진행하였다. 또한 데이터의 특성을 파악하여 특성에 가장 적합한 분석방법을 사용하고자 하였다. 셋째, 데이터 정제과정을 거친 데이터를 IBM SPSS Modeler 16.0 프로그램을 통해 데이터마이닝 기법인 의사결정

나무와 인공신경망, 로지스틱 회귀분석을 사용하여 종속변수에 가장 큰 영향을 끼치는 독립변수를 선정하였다. 마지막으로 변수들이 끼치는 영향력에 대해 알아보기 위해 추가적으로 회귀분석을 실시하였다. 회귀분석은 R 프로그래밍 기법을 사용하였으며 종속변수에 가장 큰 영향을 끼친 상위 독립변수들과 종속변수 사이의 상관관계에 대해 분석하였다.

이러한 네 단계의 절차를 거쳐 결론을 도출하고 이를 바탕으로 관광산업 활성화 방안을 제시하고자 한다.

3.2 데이터 전 처리

본 연구는 한국을 방문하는 외래 관광객의 관광과정, 관광활동, 관광평가 등 관광실태를 파악

Procedure	Main tasks	Research method and analysis
Step 1	Collecting data	Collecting data on "Survey on the actual situation of foreign tourists from 2013 to 2015" surveyed by the Korea Culture & Tourism Research Institute.
Step 2	Data Cleaning	<ul style="list-style-type: none"> • Data integration • Remove unnecessary variables for analysis • Creating derived variables • Variable transformation
Step 3	Selecting Independent Variables Using Data Mining Techniques	<ul style="list-style-type: none"> • Selection of analysis techniques (decision tree, artificial neural network, logistic regression analysis) • Selection of dependent variables (Satisfaction, Recommendation intention, Revisit intention) • Selection of the model with the highest accuracy • Select 7 top variables of the model with the highest accuracy
Step 4	Independent variable influence verification through R programming	<ul style="list-style-type: none"> • Validation of the seven variables that are the top variables of the model with the highest accuracy through r programming • In addition, logistic regression analysis for each dependent variable
Step 5	Conclusion	<ul style="list-style-type: none"> • Influence analysis of each variable by dependent variable • Identify the factors that have the greatest effect on each dependent variable

<Figure 1> Research process

하기 위하여 2013년부터 2015년까지 한국을 방문한 외래 관광객을 대상으로 실시한 “외래 관광 실태조사” 데이터를 활용하였다. 국가(지역)은 일본, 중국, 홍콩, 싱가포르, 대만, 태국, 말레이시아, 호주, 미국, 캐나다, 영국, 독일, 프랑스, 러시아, 중동, 인도, 기타로 구분하였다. 조사대상은 한국을 방문한 후 출국하는 외래 관광객으로 만 15세 이상이고 최소 체류기간은 1일 이상, 최대 1년 이하로 체류하는 사람이다. 관광목적 방한이 아닌 기장, 승무원, 군인은 제외되었고 환승을 위해 한국에서 숙박만 하는 경우도 제외되었다. 2인 이상 동반 시 1명만 조사 대상에 포함되었고 10인 이상 동반자는 2명까지 조사되었다. 단체여행은 동일 패키지 상품에서 4명까지 조사에 포함하여 공정성을 높였다. 조사규모는 매달 약 1,000명을 대상으로 주로 인천국제공항, 김포국제공항, 제주국제공항, 인천항 국제여객터미널, 부산항 국제여객터미널에서 조사되었으며 조사기간은 1년 단위로 진행되었다 (관광정보지식시스템 <http://www.tour.go.kr>).

데이터 정제과정으로는 먼저 “2013년~2015년 외래 관광객 실태조사” 데이터를 통합하여 한국 방문 직전 방문 국가, 여행 전 방문 검토국가, 한국여행 결정 시점, 쇼핑품목, 쇼핑장소와 같은 본 분석에서 불필요한 변수들을 제거하였다. 그리고 분석의 정확도를 높이하고자 5점 척도로 된 설문조항을 1~3번은 불만족으로, 4~5번은 만족으로 변수를 변형하였다. 설문조사 특성상 보통이다(3번)는 부정적 의견으로 정의하였다. 또한 체류기간은 평균값인 6.6일을 기준으로 1~4일을 단기, 5~10일을 중기, 10일 이상을 장기여행객으로 분류하여 분석을 진행하였다. 최종적으로 데이터 정제를 통해 29개의 변수를 선정하였고, 약 36,000개의 데이터를 사용하였다. 분석에 활용한

변수는 <Table 1>과 같다.

<Table 1> Variable overview

Analytical variable	Explanation
Gender	성별
Edu	학력별
Age	연령별
Job	직업별
Nation	거주국별
Purpose	방한 목적
Motif	여행 동기
Stay	체류기간
Visit Type	여행 형태
Partner	동반현황
N Partner	동반자수
Activity	방문활동내역
N Visit	방문지역수
B.Attraction	좋았던 방문지
Appeal	관광지매력도
B.Activity	제일 좋았던 활동
N.Attraction	방문 관광지수
S.Guide	관광안내서비스만족도
Information	여행정보수집경로
S.Transport	대중교통만족도
Indi Transport	개별교통비
Group Transport	단체교통비
Expense	1인당 총 경비
Revisit	재방문여부
S.Food	음식만족도
S.Stay	숙박만족도
Satisfaction	전반적인 만족도
Revisit Intention	재방문 의사
Recommend Intention	추천 의사

본 연구는 궁극적으로 한국관광 관광객의 전반적인 실태를 빅 데이터를 통해 파악해 봄으로써 한국관광산업의 활성화를 도모하는 것이다. 이에 관광객들의 전반적인 만족도를 종속변수로

선정하였다. 또한, 재방문 의사와 타인에게 추천하고자 하는 의사도 한국 관광에 대한 만족을 파악하는데 중요한 척도라고 생각하여 전반적인 만족도, 재방문 의사, 추천의사 총 세 가지의 변수를 종속변수로 선정하였다.

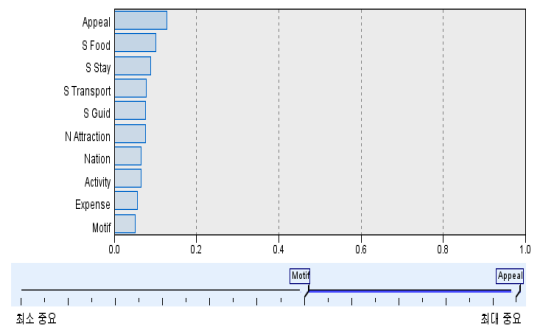
4. 데이터마이닝

본 연구에서는 변수 중 전반적 만족도, 추천의사, 재방문의사를 종속변수로 선정하고 각각의 종속변수에 가장 큰 영향을 미치는 상위변수를 7개씩 도출하였다. 분석 도구로는 IBM SPSS Modeler16.0을 사용하여 데이터마이닝 기법 중 의사결정나무, 인공신경망, 로지스틱 회귀분석 기법을 사용하였다. 의사결정나무에서는 C5.0, CART, CHAID, QUEST 등 4가지 알고리즘을 사용하여 분석을 진행하였고 이중 정확도가 가장 높은 결과를 채택하였다.

4.1 전반적 만족도 분석

먼저 전반적 만족도(Satisfaction)를 종속변수로 선정하고 분석을 진행하였다. 의사결정나무, 인공신경망, 로지스틱 회귀분석의 데이터마이닝 기법들을 적용하여 정확도를 측정한 결과 C5.0이 가장 높은 정확도를 나타냈으며 그 다음으로 CART, CHAID, QUEST 인공신경망, 로지스틱 회귀분석 순으로 높은 정확도를 보였다. 이 중 가장 정확도가 높은 결과는 C5.0으로 90.52%의 정확성을 보여주었다. C5.0을 통해 본 분석결과로 종속변수인 전반적 만족도(Satisfaction)에 가장 큰 영향을 미치는 상위 변수는 관광지매력도(Appeal), 음식만족도(S.Food), 숙박만족도(S.Stay),

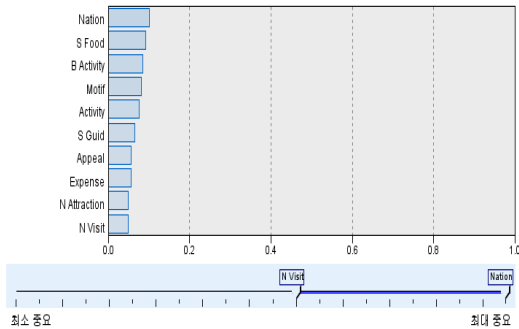
교통수단만족도(S.Transport), 관광안내서비스만족도(S.Guide), 방문관광지수(N.Attraction), 국가(Nation) 순으로 나타났다. 자세한 내용은 <Figure 2>와 같다.



<Figure 2> Predictive variable importance variable: Satisfaction

4.2 재방문의사 분석

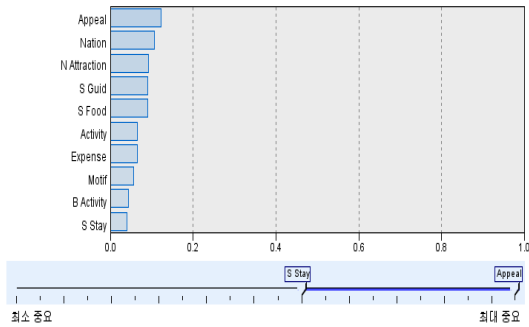
두 번째로 재방문의사(Revisit intention)를 종속변수로 선정하고 데이터마이닝 기법들을 적용하여 정확도를 측정한 결과 전반적 만족도(Satisfaction)와 동일하게 C5.0이 가장 높은 정확도를 나타냈고 다음으로 CART, CHAID, QUEST, 인공신경망의 정확도, 로지스틱 회귀분석 순이었다. 이 중 C5.0은 73.19%의 정확도로 타 분석기법에 비해 상대적으로 높은 정확성을 보여주었고 재방문의사(Revisit intention)에 가장 큰 영향을 미치는 상위 변수 7개로는 국가(Nation), 음식만족도(S.Food), 제일 좋았던 활동(B.Activity), 여행동기(Motif), 활동(Activity), 관광안내서비스만족도(S.Guide), 관광지매력도(Appeal) 순으로 나타났다. 자세한 내용은 <Figure 3>과 같다.



〈Figure 3〉 Predictive variable importance variable: Revisit intention

4.3 추천 의사 분석

마지막으로 추천 의사(Recommend Intention)를 종속변수로 선정하고 데이터마이닝 기법을 이용하여 분석한 결과 전반적 만족도(Satisfaction)와 재방문의사(Revisit intention)의 분석결과와 동일하게 C5.0이 78.71%로 타 분석 기법에 비해 높은 정확성을 보여주었다.



〈Figure 4〉 Predictive variable importance variable: Recommendation

C5.0의 분석결과로 <Figure 4>과 같이 관광지 매력도(Appeal), 국가(Nation), 방문관광지수

(N.Attraction), 관광안내서비스만족도(S.Guide), 음식만족도(S.Food), 활동(Activity), 비용(Expense) 순으로 추천의사(Recommend Intention)에 영향을 주는 것으로 나타났다.

5. 회귀분석

본 연구에서는 우선 데이터마이닝적인 접근으로 종속변수에 어떤 변수가 가장 중요한 영향을 끼치는지 알아보았고 이중 가장 큰 영향을 끼치는 상위변수들을 도출하여 R 프로그래밍 기법으로 로지스틱 회귀분석을 추가적으로 진행하였다. 종속변수에 가장 큰 영향을 미치는 상위 7개의 독립변수들을 이용해 다시 한 번 로지스틱 회귀분석을 진행함으로써 각각의 독립변수들이 종속변수에 끼치는 영향성에 대해 더욱 심도 깊은 분석을 하였다.

5.1 만족도(Satisfaction)에 대한 회귀분석

$$\text{Satisfaction} = \beta_0 + \beta_1 \text{Appeal} + \beta_2 \text{S.Food} + \beta_3 \text{S.Stay} + \beta_4 \text{S.Transport} + \beta_5 \text{S.Guide} + \beta_6 \text{Nation} + \beta_7 \text{N.Attraction}$$

첫 번째로 전반적 만족도(Satisfaction)에 대한 주요 상위변수로는 관광지 매력도(Appeal), 음식만족도(S.Food), 숙박만족도(S.Stay), 교통수단만족도(S.Transport), 관광안내서비스만족도(S.Guide), 방문관광지수(N.Attraction), 국가(Nation) 등 7개의 변수가 데이터마이닝 기법을 활용하여 도출되었다. 해당 변수들을 독립변수로 전반적만족도(Satisfaction)를 종속변수로 로지스틱 회귀분석을 진행한 결과 유의수준 0.001수준에서 유의한 결과로서 독립변수 관광지 매력도(Appeal)가

0.843063***, 음식만족도(S.Food)가 1.211331***, 숙박만족도(S.Stay)가 0.797701***, 교통수단만족도(S.Transport)가 0.875098***, 관광안내서비스만족도(S.Guide)가 0.862601***, 방문관광지수(N.Attraction)가 0.067609***로 유의한 결과로 나타났다(<Table 2> 참조).

그 중 β 값이 가장 큰, 즉 종속변수에 끼치는 영향이 상대적으로 큰 독립변수는 음식만족도(S.Food)로 나타났다. 또한 교통수단만족도(S.Transport), 관광안내서비스만족도(S.Guide), 관광지매력도(Appeal)도 다른 변수들에 비하여 상대적으로 큰 영향을 끼치는 독립변수임을 확인할 수 있었다. 따라서 한국여행에 대한 전반적 만족도(Satisfaction)는 음식에 대한 만족이 클수록, 관광지에 대해서 매력적으로 느낄수록 다른 경우에 비하여 상대적으로 전반적인 만족도가 크게 상승한다는 것을 확인할 수 있었다.

5.2 재방문 의사(Revisit Intention)에 대한 회귀분석

$$\begin{aligned} \text{Revisit Intention} = & \beta_0 + \beta_1 \text{Nation} + \beta_2 \text{Motif} + \beta_3 \text{Activity} \\ & + \beta_4 \text{S.Food} + \beta_5 \text{B.Activity} + \beta_6 \text{S.Guide} + \beta_7 \text{Appeal} \end{aligned}$$

두 번째로 재방문의사(Revisit intention)에 대한 중요한 상위변수로는 국가(Nation), 여행동기(Motif), 활동(Activity), 음식만족도(S.Food), 제일 좋았던 활동(B.Activity), 관광안내서비스만족도(S.Guide), 관광지매력도(Appeal) 등 7개의 변수가 데이터마이닝 기법을 통하여 도출되었다. 해당 변수들을 독립변수로 재방문의사(Revisit intention)를 종속변수로 로지스틱 회귀분석을 진행한 결과 유의수준 0.001수준에서 유의한 결과

로 독립변수 국가(Nation)가 대만(-0.559962***), 독일(-0.501731***), 러시아(-0.764134***), 미국(-0.457782***), 중국(-0.359576***), 캐나다(-0.679748***), 태국(-0.674446***), 프랑스(-0.483579***)인 경우와 독립변수 여행동기(Motif)가 한류(1.23679***), 활동(Activity)가 업무수행(0.390751***), 제일 좋았던 활동(B.Activity)이 쇼핑(0.258802***), 식도락(0.33257***), 업무수행(0.416709***)인 경우 유의한 결과로 나타났다. 또한, 음식만족도(S.Food)가 0.908337***, 관광안내서비스만족도(S.Guide)가 0.467111***, 관광지매력도(Appeal)가 0.561776***로 유의한 결과로 나타났다(<Table 3> 참조).

그 중 β 값이 가장 큰, 즉 종속변수에 끼치는 영향이 상대적으로 큰 독립변수는 여행동기(Motif)로 나타났으며 여행동기(Motif)가 한류인 경우 다른 독립변수들에 비하여 재방문의사(Revisit intention)에 커다란 긍정적인 영향을 끼친다는 것을 확인할 수 있었다. 또한 전반적만족도(Satisfaction)와 동일하게 음식만족도(S.Food)가 다른 독립변수들에 비하여 β 값인 0.908337***만큼 상대적으로 큰 영향을 끼치는 독립변수임을 확인할 수 있었다.

유의수준 0.01수준에서는 여행동기(Motif)중 여가시간(-0.441135**)이, 활동(Activity)중엔 카지노 (0.838203**), 공연,행사,축제(0.463653**), 식도락(0.259033**), 쇼핑(0.225037**)이, 제일 좋았던 활동(B.Activity)엔 의료관광(0.865005**), 시티투어버스(0.272757**) 등이 유의한 것으로 나타났다.

따라서 한국여행에 대한 재방문의사(Revisit intention)가 높은 경우에는 다양한 여행동기

(Motif) 중 한류의 목적으로 방문한 경우 다른 경우에 비해 상대적으로 재방문의사가 높게 나타났으며 이에 따라 한류를 목적으로 방문한 관광객들의 높은 재방문의사(Revisit intention)를 실질적인 재방문으로 이어지게 할 수 있는 정책이 필요할 것으로 보인다. 또한 전반적만족도(Satisfaction)와 동일하게 음식만족도(S.Food)가 높을수록 재방문의사(Revisit intention)가 높아지기 때문에 음식만족도(S.Food)를 높일 수 있는 방안 모색도 함께 필요할 것으로 보인다. 기타 카지노와 의료관광분야에 대한 수요가 높음을 알 수 있는데 이러한 분야의 특성화에 관심을 기울일 필요가 있다.

5.3 추천의사(Recommend Intention)에 대한 회귀분석

$$\begin{aligned} \text{Recommend Intention} = & \beta_0 + \beta_1 \text{Nation} + \beta_2 \text{Appeal} + \beta_3 \text{N.Attraction} \\ & + \beta_4 \text{S.Guide} + \beta_5 \text{Activity} + \beta_6 \text{S.Food} + \beta_7 \text{Expense} \end{aligned}$$

마지막으로 3가지 종속변수 중 데이터마이닝 기법을 통한 추천의사(Recommend Intention)에 대한 중요한 상위변수로는 국가(Nation), 관광지 매력도(Appeal), 방문관광지수(N.Attraction), 음식만족도(S.Food), 활동(Activity), 관광안내서비스만족도(S.Guide), 비용(Expense) 등 7개의 변수가 데이터마이닝 기법을 활용하여 도출되었다. 해당 변수들을 독립변수로 추천의사(Recommend Intention)를 종속변수로 로지스틱 회귀분석을 진행한 결과 유의수준 0.001수준에서의 유의한 결과로는 독립변수 국가(Nation)가 대만(-1.042***), 독일(-0.939***), 러시아(-0.9785***), 일본(-1.047***), 중국(-0.7345***), 태국

(-0.6878***), 프랑스(-0.6347***), 홍콩(-0.6115***)인 경우와 독립변수 관광지매력도(Appeal)이 0.934***, 방문관광지수(N.Attraction)이 0.03217***, 관광안내서비스만족도(S.Guide)가 0.6478***, 음식만족도(S.Food)가 1.058***로 유의한 결과로 나타났으며 독립변수 활동(Activity) 중 휴양, 휴식이 -0.6178***로 유의한 결과로 나타났다(<Table 4> 참조).

그 중 β 값이 가장 큰, 즉 종속변수에 끼치는 영향이 상대적으로 큰 독립변수로는 국가(Nation) 중 대만, 독일, 러시아, 일본인 경우였으며 해당 국가가 아닌 경우에 비하여 추천의사(Recommend Intention)에 상대적으로 β 값인 -1.042***, -0.939***, -0.9785***, -1.047***만큼 부정적인 영향을 끼친다는 것을 확인할 수 있었다.

또한 전반적만족도(Satisfaction)와 동일하게 관광지매력도(Appeal)와 음식만족도(S.Food)가 각각 만족하지 않은 경우에 비하여 만족한 경우에 0.934***, 1.058***만큼 상대적으로 다른 독립변수들에 비하여 종속변수인 추천의사(Recommend Intention)에 큰 영향을 끼치는 것을 확인할 수 있었다.

따라서 한국여행에 대한 추천의사(Recommend Intention)가 높은 국가보다는 상대적으로 추천의사(Recommend Intention)가 낮은 국가들에 대해서 그에 대한 원인 분석과 방안마련이 필요할 것으로 보이며, 관광지매력도(Appeal)와 음식만족도(S.Food)를 높일 수 있는 방안도 계속해서 모색하는 것이 중요할 것으로 사료된다.

본 연구에서 세 가지 종속변수 추천의사(Recommend Intention), 재방문의사(Revisit intention), 전반적만족도(Satisfaction)에 공통적

으로 가장 큰 영향을 끼치는 독립변수는 음식만족도(S.Food)와 관광지 매력도(Appeal)로 분석되었다.

즉, 다른 독립변수들인 여행목적이나 여행비용, 방문관광지 수 등에 비하여 음식에 대한 만족도가 높을수록 또한 관광지의 매력도가 상승할수록 여행에 대한 만족도가 상승할 뿐만 아니라 한국여행을 추천하고자 하는 의사, 재방문하고자 하는 의사가 높게 나타나는 것으로 나타났다.

뿐만 아니라 전반적만족도(Satisfaction)와 재방문의사(Revisit intention) 그리고 추천의사(Recommend Intention)에 국가(Nation)가 공통적으로 높은 영향을 미치고 있는 것으로 확인되었으며, 국가에 따라 한국여행에 대한 만족도가 떨어지는지 이유가 무엇인지 추가적으로 그 원인을 파악하고 그에 대한 방안을 모색하는 연구도 필요할 것으로 사료된다.

6. 결론

전 세계적으로 관광산업이 차지하는 GDP의 비중이 증가하고 외래 관광객의 규모와 그 관광산업의 중요도가 날로 높아짐에 따라 관광산업 활성화를 위한 국가 간의 경쟁이 더욱더 치열해지고 있다. 본 연구는 한국문화관광연구원에서 조사한 2013년부터 2015년까지의 외래 관광객 실태조사를 데이터를 통합하여 약 36,000개의 빅 데이터를 IBM SPSS Modeler 16.0을 활용하여 의사결정나무, 인공신경망, 로지스틱회귀분석을 진행하였으며 해당 분석을 통해 얻은 외래 관광객의 여행 만족도에 가장 큰 영향을 끼치는 상위변수들을 위주로 추가적으로 로지스틱 회귀

분석을 진행하여 그 영향력에 대해 심층분석을 하였다.

우선 데이터마이닝 기법을 통해 각 종속변수별 종속변수에 가장 큰 영향을 미치는 상위변수를 7개씩 도출하였다. 첫 번째로 전반적 만족도로 데이터마이닝을 시행하였을 때 가장 높은 정확도를 보인 모델은 의사결정나무의 C5.0이었으며 90.52%로 가장 높은 정확도를 보여주고 있었다. C5.0의 분석결과 전반적 만족도에 가장 큰 영향을 미치는 상위변수는 관광지 매력도(Appeal), 음식만족도(S.Food), 숙박만족도(S.Stay), 교통수단만족도(S.Transport), 관광안내서비스만족도(S.Guide), 방문관광지수(N.Attraction), 국가(Nation) 순으로 나타났다. 상위변수에 대한 각각의 영향력을 파악하기 위하여 상위 변수 7개로 R프로그래밍을 통해 로지스틱 회귀분석을 추가적으로 시행한 결과 음식만족도(S.Food)가 종속변수에 가장 큰 영향을 끼치는 변수로 나타났으며 만족인 경우 불만족인 경우에 비하여 전반적 만족도에 긍정적인 영향을 끼친다는 것을 확인할 수 있었다. 또한 관광지 매력도(Appeal)도 음식만족도 다음으로 전반적 만족도에 큰 영향을 끼치는 것으로 나타났다.

두 번째로 재방문 의사를 종속변수로 데이터마이닝을 시행한 결과 가장 높은 정확도를 보인 모델은 의사결정나무의 C5.0으로 나타났으며 73.19%로 타 모델에 비하여 상대적으로 높은 정확도를 보였다. 재방문의사에 가장 큰 영향을 미치는 상위 변수는 국가(Nation), 음식만족도(S.Food), 제일 좋았던 활동(B.Activity), 여행동기(Motif), 활동(Activity), 관광안내서비스만족도(S.Guide), 관광지 매력도(Appeal) 순으로 나타났다. 상위 변수 7개로 R프로그래밍을 통해 로지스틱 회귀분석을 추가적으로 시행한 결과 여행동

기(Motif)가 한류인 경우 다른 여행동기에 비하여 긍정적인 영향을 끼친다는 것을 확인할 수 있었으며, 전반적 만족도와 동일하게 음식만족도(S.Food)가 종속변수에 큰 영향을 끼치는 변수로 나타났다.

마지막으로 추천 의사를 종속변수로 데이터마닝을 시행한 결과 가장 높은 정확도를 보인 모델은 동일하게 의사결정나무의 C5.0으로 78.71%의 정확도를 보였다. 추천 의사에 가장 큰 영향을 미치는 상위 변수는 관광지매력도(Appeal), 국가(Nation), 방문관광지수(N.Attraction), 관광안내서비스만족도(S.Guide), 음식만족도(S.Food), 활동(Activity), 비용(Expense) 순으로 나타났다. 상위 변수 7개로 R프로그래밍을 통해 로지스틱 회귀분석을 추가적으로 시행한 결과 국가(Nation)가 종속변수에 가장 큰 영향을 끼치는 변수로 나타났으며 대만, 독일, 러시아, 일본이 다른 국가에 비하여 상대적으로 추천 의사에 부정적인 영향을 가진다는 것을 확인할 수 있었다. 또한 전반적 만족도와 동일하게 관광지매력도(Appeal)와 음식만족도(S.Food)가 각각 종속변수에 큰 영향을 끼치는 것을 확인할 수 있었다.

빅 데이터 분석 기법들을 통해 외래 관광객들의 만족도에 가장 큰 영향을 미치는 요인을 분석한 결과 각각의 종속변수에 공통적으로 가장 큰 영향을 주는 변수는 음식만족도와 관광지 매력도였다. 해당 두 가지 요인은 여행비용이나 여행 목적 등 다른 요인들에 비해 전체 외래 관광객들의 여행 만족도에 상대적으로 큰 영향을 끼친다는 사실을 파악하였으며 해당 요인을 참고하여 관광산업에 응용하면 한국 관광 활성화에 기여할 수 있을 것으로 기대한다. 본 연구를 바탕으로 향후 기업 및 국가차원에서 한국 관광발전에

기여할 수 있는 활성화 방안으로 음식만족도를 높이기 위한 맛집 또는 음식 맛 평가에 대한 여행객들끼리의 공유 플랫폼 구축 등을 고려해 볼 수 있을 것이다. 또한, 관광지매력도를 높이기 위해 실질적으로 외래 관광객이 매력을 느낄 수 있는 한국적인 관광지 마련에 힘써야 할 것이다. 의료관광이나 카지노와 같은 특화 서비스를 육성 하거나 스마트폰 같은 신기술을 활용(Bae, 2010) 하여 국가적 차원에서는 이미 개발되어 있는 지역관광인프라를 최대한 활용할 수 있는 방안을 마련하여 지역관광 활성화에도 힘써야 할 것이다.

본 연구는 설문조사 특성상 긍정적인 답변에 치우친 결과를 얻을 수밖에 없다는 한계점이 존재한다. 또한, 3개년도의 외래 관광객 실태를 바탕으로 연구를 진행하였으나 앞으로 좀 더 장기간의 관광 트렌드를 연구해 볼 필요가 있다.

참고문헌(References)

- Bae, K.-Y., , "Self-Tour Service Technology based on a Smartphone", *Journal of Intelligence and Information Systems*, Vol. 16, No. 4 (2010), 147~157.
- Byun, C. B., "The oversea backpackers' travel experiences : Based on grounded theory," *International Journal of Tourism and Hospitality Research*, Vol.27, No.2(2012), 239-262.
- Choi, J. H., S. T. Han, H. C. Kang and E. S. Kim, "Data Mining Decision Tree Analysis Using Answer Tree", *SPSS ACADEMY*, (1998), 17-23.
- Choi, Y. J. and D. H. Park, "Development of

- Yókè Mining System with Yókè's Travel Demand and Insight Based on Web Search Traffic Information”, *Journal of Intelligence and Information Systems*, Vol. 23, No. 3(2017), 155-175.
- Do, H. Y., A. J. Lee and C. W. Lee, "The Study of Restaurant Service Quality, Customer Satisfaction and Nationality by Big-Data Analysis", *Journal of Foodservice Management Society of Korea*, Vol.16, No.6 (2013), 49-70.
- Hwang, R., "A Study on the Trend Analysis of Chinese Tourists: Focusing on Survey of Foreign Tourists", Master Dissertation, Kyung Gi University, (2015)
- Hunt, H. K., "Conceptualization and measurement of consumer satisfaction and dissatisfaction”, *Proceedings of Marketing Science Institute*(No. 77-103), (1977).
- Kim, D. W., "Big Data Use Cases of the Sector", *The Journal of Management*, Vol. 34 (2013), 39-52.
- Kim, M. H., S. W. Choo and Y. H. Hwang, "Research Articles : Factors Influencing Chinese Travelers` Multi-destination Trip in Korea: Capital Area versus Other Areas", *Journal of Tourism Sciences*, Vol.35, No.9(2011), 283-296.
- Kim, C. O. and Y. K. Song, "The Effect of Tourist Experience on Tourist Satisfaction and Quality of Life : Case of Non-institutionalized Domestic Tourists", *International Journal of Tourism and Hospitality Research*, Vol.21, No.4(2007), 23-42.
- Kim B. H., "Study on the Relationship and Trade-offs between Tourism and Aviation Policy”, *Journal of Tourism Enhancement*, Vol.2, No.1(2014), 51-70.
- Kim, G. G., C. W. Yeo, Y. C. Kim, "A Study on Chinese Visitors' Travel product selection attribute and Satisfaction”, *Korean Journal of Tourism Research*, Vol. 22, No.1(2007), 133-153.
- Ko, D. W., "The identity of leisure · tourism psychology”, *Proceedings of Korea Academic Society of Tourism Mngement*, (2006), 20-23.
- Kong, H. S. and E. J. Song, "A Study on Hotel CRM(Customer Relationship Management) using Big Data and Security”, *Journal of Information and Security*, Vol.13, No.4 (2013). 69-75.
- Lee J. M., M. N. Ryu and G. G. Lim, "A Study on the Revitalization of Tourism Industry through Big Data Analysis”, *Proceedings of the Korea Intelligence Information Systems Society*, (2016), 73-73.
- Lee, Y. J., H. K. Jeong. and S. H. Boo, "Evaluation of the Publicity on eTourism Information: Focused on the Portal Site 'N'", *Journal of Tourism & Leisure Research*, Vol. 22, No. 3(2010), 389-408.
- Lee, Y. J. and J. H. Yoon, "A Study on Utilizing SNS Big Data in the Tourism Studies : Based on an Analysis of Key Words for Tourism Information Search", *International Journal of Tourism and Hospitality Research*, Vol. 28, No. 3(2014), 5-14.
- Lee, J. H., "The Impact of Tourism Images on Satisfaction, Revisit Intention, Willingness to Recommend -Centered on the Foreign Tourists` Evaluation of Busan Tourism Images-", *Journal of Hospitality and Tourism*

- Studies*, Vol.31, (2008), 183-198.
- Lee, J. K. and E. H. Kim, "The Study on the Tourism Satisfaction and Revisit of Tour Experience toward Foreign Tourist - Focused on Bukchon Guesthouse -", *International Journal of Tourism and Hospitality Research*, Vol.29, No.1(2014), 179-199.
- Lee S. H., D. H. Park and M. S. Kim, "A Study of Perceptions of Jeju-Island Tourism using Social Media Big Data Analysis: Before and after the Outbreak of Ban on Korean entertainer"s", *International Journal of Tourism and Hospitality Research*, Vol. 30, No.2(2018), 55-75.
- Oh, I. K., T, S, Lee and C. N. Chon, "A Study on Awareness of Korea Tourism through Big Data Analysis", *Journal of Tourism Sciences*, Vol.39, No.10, (2015), 107-126.
- Oh, J. K., " Market Segmentation Based on the Motivation of Marine Tourism and the effect of Marine Tour Activities on Tour Satisfaction and Image - Focused on Busan Area", *International Journal of Tourism and Hospitality Research*, Vol.27, No.1(2012), 17-36.
- Park, B. E. and G. G. Lim, "A Study on the Impact Factors of Contents Diffusion in Youtube using Integrated Content Network Analysis", *Journal of Intelligence and Information Systems*, Vol. 21, No. 3(2015), 41-58.
- Park I. G. and K. H. Kang, "A Study on the Selection of Tourist Destinations of Chinese Tourists", *Korea Journal of Tourism and Hospitality Research*, Vol. 23, No.4(2009), 67-80.
- Park, Y. K., "Effects of Cruise Tourism Product Attributes on Tourist's Satisfaction -Centered around Korean Tourists Visiting Japan-", *International Journal of Tourism and Hospitality Research*, Vol.25, No.1(2010), 439-459.
- Parroco, A. M., Vaccina, D., De Cantis, S. and Ferrante, M., "Multi-destination trips & tourism statistics, Empirical evidence in sicily", *Economics*, Vol. 6, No. 44(2012), 1.
- Philip R., Big Data Analytics. *TDWI Research Fourth Quarter*, 6. (2011).
- Pizam, A., Neumann, Y., and Reichel, A., "Dimensions of tourist satisfaction with a destination area", *Annals of tourism Research*, Vol.5, No.3(1978), 314-322.
- Stankey, G. H., A strategy for the definition and management of wilderness quality. (1972).
- Tourism Knowledge Information System. "2013 Outbound Tourist Survey Source Data, 2013, Available at, <http://know.tour.go.kr>, (Downloaded 20 July, 2016).
- Tourism Knowledge Information System. "2014 Outbound Tourist Survey Source Data", 2014, Available at, <http://know.tour.go.kr>, (Downloaded 20 July, 2016).
- Tourism Knowledge Information System.. "2015 Outbound Tourist Survey Source Data", 2015, Available at, <http://know.tour.go.kr>, (Downloaded 20 July, 2016).
- Whang, R., "Analysis on the Trend of Chinese Tourists: Focusing on Survey of Foreign Tourists", Master Dissertation, Kyung Gi University, (2015).

Appendix

〈Table 2〉 Satisfaction

Coefficients:	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	-0.385456*	0.174874	-2.204	0.027511
Appeal	0.843063***	0.067566	12.478	< 2e-16
S.Food	1.211331***	0.061188	19.797	< 2e-16
S.Stay	0.797701***	0.065024	12.268	< 2e-16
S.Transport	0.875098***	0.0656	13.34	< 2e-16
S.Guid	0.862601***	0.069995	12.324	< 2e-16
Nation대만	-0.020082	0.193985	-0.104	0.917547
Nation독일	0.487188*	0.24565	1.983	0.047339
Nation러시아	0.029878	0.237758	0.126	9.00E-01
Nation말레이시아	-0.282428	0.206317	-1.369	0.17103
Nation무응답	1.04466***	0.273716	3.817	0.000135
Nation미국	0.484229*	0.210735	2.298	0.021572
Nation싱가포르	0.026184	0.221488	0.118	0.905894
Nation영국	0.610675*	0.259914	2.35	0.018797
Nation인도	0.38267	0.237974	1.608	0.107828
Nation일본	-0.1014	0.170107	-0.596	0.55111
Nation중국	-0.008748	0.173386	-0.05	0.959759
Nation중동	0.150212	0.228406	0.658	0.51076
Nation캐나다	0.980346**	0.299698	3.271	0.001071
Nation태국	-0.191562	0.197731	-0.969	0.332644
Nation프랑스	0.661818*	0.282626	2.342	0.019198
Nation호주	0.573045*	0.251622	2.277	0.022762
Nation홍콩	0.345981.	0.201403	1.718	0.085823
N.Attraction	0.067609***	0.010039	6.735	1.64E-11

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

〈Table 3〉 Revisit Intention

Coefficients:	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	-0.002057	0.177095	-0.012	0.990732
Nation대만	-0.559962***	0.10429	-5.369	7.91E-08
Nation독일	-0.501731***	0.123184	-4.073	4.64E-05
Nation러시아	-0.764134***	0.118322	-6.458	1.06E-10
Nation말레이시아	-0.093077	0.117026	-0.795	0.426408
Nation무응답	-0.309273*	0.120949	-2.557	0.010556
Nation미국	-0.457782***	0.102448	-4.468	7.88E-06
Nation싱가포르	0.116341	0.128642	0.904	0.365793
Nation영국	-0.305298*	0.125314	-2.436	0.01484
Nation인도	0.34838*	0.143145	2.434	0.014943
Nation일본	-0.149756	0.096804	-1.547	0.121864
Nation중국	-0.359576***	0.09527	-3.774	0.00016
Nation중동	0.179682	0.13894	1.293	0.195928
Nation캐나다	-0.679748***	0.115064	-5.908	3.47E-09
Nation태국	-0.674446***	0.105783	-6.376	1.82E-10
Nation프랑스	-0.483579***	0.1265	-3.823	1.32E-04
Nation호주	-0.195537	0.123703	-1.581	0.113948
Nation홍콩	-0.174276	0.109813	-1.587	0.112508
Motif기타	-0.136103	0.185279	-0.735	0.462593
Motif기후,사계절	-0.079959	0.187774	-0.426	0.670234
Motif무응답	0.365482*	0.142492	2.565	0.01032
Motif쇼핑	0.34118*	0.143842	2.372	0.017696
Motif숙박,교통	0.179812	0.298102	0.603	0.546382
Motif안전,치안	0.326106	0.231013	1.412	0.158057
Motif여가시간	-0.441135**	0.164375	-2.684	0.007281
Motif여행비용	-0.20762	0.178277	-1.165	0.244186
Motif역사,유적	0.153595	0.152782	1.005	0.31474
Motif유희,놀이	-0.066849	0.185125	-0.361	0.718024
Motif음식,미식	0.13967	0.159174	0.877	0.38023
Motif이미용서비스	0.087905	0.327022	0.269	0.78808
Motif자연풍경	0.152183	0.14302	1.064	0.2873
Motif치료	0.161538	0.287843	0.561	0.574659
Motif패션,문화	0.157231	0.163478	0.962	0.336159
Motif한류	1.23679***	0.213314	5.798	6.71E-09
Motif휴양시설	-0.084139	0.16296	-0.516	0.605633
Activity공연,행사,축제	0.463653**	0.175429	2.643	0.008218
Activity기타	0.507804*	0.214213	2.371	0.017761
Activity레포츠활동	-0.161952	0.19329	-0.838	0.402103
Activity미팅,회의,학술대회,박람회	-0.071555	0.117863	-0.607	0.543781
Activity박물관,전시관	-0.305971	0.206575	-1.481	0.138564
Activity뷰티관광	0.402975	0.291386	1.383	0.166678

Coefficients:	Estimate	Std. Error	z value	Pr(> z)
Activity쇼핑	0.225037**	0.071864	3.131	0.00174
Activity스포츠활동	0.149246	0.185837	0.803	0.421918
Activity시찰	0.082949	0.233112	0.356	0.721964
Activity시티투어버스	-0.158374	0.113888	-1.391	0.164342
Activity식도락	0.259033**	0.094181	2.75	0.005953
Activity업무수행	0.390751***	0.091608	4.265	2.00E-05
Activity연수,교육,연구	0.081394	0.120874	0.673	0.500705
Activity온천,스파	-0.100578	0.172178	-0.584	0.559119
Activity유흥,오락	-0.154673	0.134866	-1.147	0.251438
Activity의료관광	-0.196766	0.307504	-0.64	0.522251
Activity자연경관	-0.038975	0.087848	-0.444	0.657287
Activity전통문화체험	0.407547*	0.185733	2.194	0.028216
Activity직업운동	-0.064964	0.323665	-0.201	0.840922
Activity카지노	0.838203**	0.318465	2.632	0.008488
Activity테마파크	-0.000689	0.142057	-0.005	0.99613
Activity휴양,휴식	-0.297853*	0.117196	-2.541	0.011038
S.Food	0.908337***	0.037018	24.538	< 2e-16
B.Activity공연,행사,축제	0.234163.	0.121788	1.923	0.054517
B.Activity기타	0.301119	0.204694	1.471	0.141271
B.Activity레포츠활동	0.005829	0.150748	0.039	0.969154
B.Activity미팅,회의,학술대회,박람회	0.139468	0.123486	1.129	0.258721
B.Activity박물관,전시관	0.337739.	0.195495	1.728	0.084059
B.Activity뷰티관광	0.142284	0.185265	0.768	0.442486
B.Activity쇼핑	0.258802***	0.067425	3.838	0.000124
B.Activity스포츠활동	0.154775	0.142734	1.084	0.278208
B.Activity시찰	0.56913*	0.26692	2.132	0.03299
B.Activity시티투어버스	0.272757**	0.103731	2.629	0.008552
B.Activity식도락	0.33257***	0.078282	4.248	2.15E-05
B.Activity업무수행	0.416709***	0.09154	4.552	5.31E-06
B.Activity연수,교육,연구	0.23158	0.141164	1.641	0.100901
B.Activity온천,스파	-0.095032	0.128366	-0.74	0.459105
B.Activity유흥,오락	0.258967*	0.109174	2.372	0.017689
B.Activity의료관광	0.865005**	0.303224	2.853	0.004335
B.Activity자연경관	0.032794	0.077285	0.424	0.671327
B.Activity전통문화체험	-0.062488	0.118468	-0.527	0.597867
B.Activity직업운동	-0.17823	0.318099	-0.56	0.575276
B.Activity카지노	0.052187	0.163358	0.319	0.749375
B.Activity테마파크	0.22165*	0.110396	2.008	0.044667
B.Activity휴양,휴식	0.1213	0.121104	1.002	0.316528
S.Guid	0.467111***	0.0387	12.07	< 2e-16
Appeal	0.561776***	0.042064	13.355	< 2e-16

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Table 4) Recommend Intention

Coefficients:	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	0.7754***	0.1524	5.087	0.00000364
Nation대만	-1.042***	0.1369	-7.614	2.65E-14
Nation독일	-0.939***	0.1486	-6.317	2.66E-10
Nation러시아	-0.9785***	0.1549	-6.316	2.69E-10
Nation말레이시아	-0.2776.	0.1587	-1.749	0.08032
Nation무응답	-0.4111**	0.1552	-2.65	0.00806
Nation미국	-0.3803**	0.1399	-2.719	0.00655
Nation싱가포르	-0.1556	0.1648	-0.944	0.34506
Nation영국	-0.2836.	0.1616	-1.755	0.07932
Nation인도	0.006226	0.1672	0.037	0.97031
Nation일본	-1.047***	0.1263	-8.291	< 2e-16
Nation중국	-0.7345***	0.1287	-5.706	1.15E-08
Nation중동	-0.1788	0.1691	-1.058	0.29028
Nation캐나다	-0.4527**	0.1612	-2.808	0.00498
Nation태국	-0.6878***	0.1439	-4.78	0.00000176
Nation프랑스	-0.6347***	0.1604	-3.958	0.0000756
Nation호주	-0.2774.	0.1613	-1.72	0.08548
Nation홍콩	-0.6115***	0.1418	-4.311	0.0000162
Appeal	0.934***	0.04422	21.119	< 2e-16
N.Attraction	0.03217***	0.005821	5.526	3.28E-08
S.Guid	0.6478***	0.04348	14.899	< 2e-16
Activity공연,행사,축제	0.4122**	0.155	2.659	0.00783
Activity기타	0.2049	0.1754	1.168	0.24283
Activity레포츠활동	0.01979	0.2613	0.076	0.93962
Activity미팅,회의,학술대회,박람회	-0.256*	0.1135	-2.256	0.02406
Activity박물관,전시관	0.1266	0.3171	0.399	0.68965
Activity뷰티관광	0.8367*	0.3659	2.287	0.0222
Activity쇼핑	0.06925	0.08847	0.783	0.43375
Activity스포츠활동	-0.3297.	0.1834	-1.798	0.07223
Activity시찰	-0.2184	0.2077	-1.051	0.29322
Activity시티투어버스	-0.01047	0.1434	-0.073	0.9418
Activity식도락	0.1795	0.1125	1.596	0.11048
Activity업무수행	-0.2596**	0.09294	-2.793	0.00522
Activity연수,교육,연구	-0.009356	0.1299	-0.072	0.94258
Activity온천,스파	-0.1127	0.2104	-0.536	0.59213
Activity유흥,오락	0.02714	0.169	0.161	0.87243
Activity의료관광	-0.006827	0.2264	-0.03	0.97594
Activity자연경관	-0.1282	0.109	-1.176	0.23955
Activity전통문화체험	-0.2002	0.2207	-0.907	0.36432
Activity직업운동	-0.6605*	0.285	-2.317	0.02048
Activity카지노	-0.1132	0.268	-0.423	0.67263
Activity테마파크	-0.09229	0.1735	-0.532	0.59478
Activity휴양,휴식	-0.6178***	0.1311	-4.713	0.00000244
S.Food	1.058***	0.04105	25.777	< 2e-16
Expense	0.00001823	0.000014	1.302	0.19284

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Abstract

A Study on the Revitalization of Tourism Industry through Big Data Analysis

Jungmi Lee* · Meina Liu* · Gyoo Gun Lim**

Korea is currently accumulating a large amount of data in public institutions based on the public data open policy and the "Government 3.0". Especially, a lot of data is accumulated in the tourism field. However, the academic discussions utilizing the tourism data are still limited. Moreover, the openness of the data of restaurants, hotels, and online tourism information, and how to use SNS Big Data in tourism are still limited. Therefore, utilization through tourism big data analysis is still low.

In this paper, we tried to analyze influencing factors on foreign tourists' satisfaction in Korea through numerical data using data mining technique and R programming technique. In this study, we tried to find ways to revitalize the tourism industry by analyzing about 36,000 big data of the "Survey on the actual situation of foreign tourists from 2013 to 2015" surveyed by the Korea Culture & Tourism Research Institute. To do this, we analyzed the factors that have high influence on the 'Satisfaction', 'Revisit intention', and 'Recommendation' variables of foreign tourists. Furthermore, we analyzed the practical influences of the variables that are mentioned above.

As a procedure of this study, we first integrated survey data of foreign tourists conducted by Korea Culture & Tourism Research Institute, which is stored in the tourist information system from 2013 to 2015, and eliminate unnecessary variables that are inconsistent with the research purpose among the integrated data. Some variables were modified to improve the accuracy of the analysis. And we analyzed the factors affecting the dependent variables by using data-mining methods: decision tree(C5.0, CART, CHAID, QUEST), artificial neural network, and logistic regression analysis of SPSS IBM Modeler 16.0. The seven variables that have the greatest effect on each dependent variable were derived.

As a result of data analysis, it was found that seven major variables influencing 'overall satisfaction' were sightseeing spot attraction, food satisfaction, accommodation satisfaction, traffic satisfaction, guide

* Business School of Hanyang University

** Corresponding Author: Gyoo Gun Lim

School of Business, Hanyang University

222, Wangsimni-ro, Seongdong-gu, Seoul, 04763

Tel: +82-2-2220-2593, Fax: +82-2-2220-1169, E-mail: ggseer@gmail.com

service satisfaction, number of visiting places, and country. Variables that had a great influence appeared food satisfaction and sightseeing spot attraction. The seven variables that had the greatest influence on 'revisit intention' were the country, travel motivation, activity, food satisfaction, best activity, guide service satisfaction and sightseeing spot attraction. The most influential variables were food satisfaction and travel motivation for Korean style. Lastly, the seven variables that have the greatest influence on the 'recommendation intention' were the country, sightseeing spot attraction, number of visiting places, food satisfaction, activity, tour guide service satisfaction and cost. And then the variables that had the greatest influence were the country, sightseeing spot attraction, and food satisfaction.

In addition, in order to grasp the influence of each independent variables more deeply, we used R programming to identify the influence of independent variables. As a result, it was found that the food satisfaction and sightseeing spot attraction were higher than other variables in overall satisfaction and had a greater effect than other influential variables. Revisit intention had a higher β value in the travel motive as the purpose of Korean Wave than other variables. It will be necessary to have a policy that will lead to a substantial revisit of tourists by enhancing tourist attractions for the purpose of Korean Wave. Lastly, the recommendation had the same result of satisfaction as the sightseeing spot attraction and food satisfaction have higher β value than other variables.

From this analysis, we found that 'food satisfaction' and 'sightseeing spot attraction' variables were the common factors to influence three dependent variables that are mentioned above('Overall satisfaction', 'Revisit intention' and 'Recommendation'), and that those factors affected the satisfaction of travel in Korea significantly. The purpose of this study is to examine how to activate foreign tourists in Korea through big data analysis. It is expected to be used as basic data for analyzing tourism data and establishing effective tourism policy. It is expected to be used as a material to establish an activation plan that can contribute to tourism development in Korea in the future.

Key Words : Big Data, Artificial Neural Network, Decision Tree, Logistic Regression Analysis,
R Programming

Received : November 6, 2017 Revised : April 11, 2018 Accepted : June 5, 2018

Publication Type : Regular Paper Corresponding Author : Gyoo Gun Lim

저 자 소개



Jungmi Lee

Jungmi Lee is a graduate student at Business School of Hanyang University. She is majoring Business Analytics. Her interesting research areas are big data analysis, intelligent systems.



Meina Liu

Meina Liu is a graduate student at Business School of Hanyang University. She is majoring Management Information Systems. Her interesting research areas are big data analysis, intelligent systems, and innovative business models.



Gyoo Gun Lim

Professor Gyoo Gun Lim is currently a Professor of MIS at School of Business of Hanyang University. He received his Ph.D. in Management Engineering from Korea Advanced Institute of Science and Technology (KAIST) in 2001. He received an award from Korea Ministry of Information and Communication for his contribution to Korea SW industry in 2007 and an award from Korea Ministry of Knowledge Economy for his contribution to Korea IT innovation in 2009. His current research interests include IT service, innovative business models, e-business, intelligent information & knowledge management, and etc.