

DEA를 적용한 문화관광축제의 효율성 등급 예측모형*

김은미** · 홍태호***

〈목 차〉

I. 서론	IV. 분석방법 및 결과
II. 이론적 배경	4.1 데이터
2.1 문화관광축제	4.2 문화관광축제의 효율성 분석
2.2 효율성 모형	4.3 기계학습을 이용한 축제 효율성 등급
2.3 다분류 모형	예측
III. 연구 프레임워크	V. 결론 및 향후 연구과제
	참고문헌
	<Abstract>

I. 서론

지역축제는 지역의 관광산업 발전과 경제 활성화를 위해 중요한 역할을 하고 있다. 문화체육관광부는 문화관광축제 지원 사업을 통해 국내의 전통문화와 독특한 주제를 바탕으로 한 지역축제 중 관광 상품성이 큰 축제를 대상으로 1995년부터 지속적으로 지원해오고 있다. 문화관광축제의 선정은 각 광역자치단체에서 추천한 축제에 대해 관광·축제 관련분야의 전문가들로 구성된 선정위원회에 의해 이루어진다. 매년 약 40개의 축제가 문화관광축제로 선

정되고 있으며 선정된 축제는 대표축제, 최우수축제, 우수축제, 유망축제로 분류하여 경쟁력 있는 지역축제로 성장할 수 있도록 지원하고 있다(문화체육관광부, 2018).

문화체육관광부는 문화관광축제로 선정된 축제만을 대상으로 매년 문화관광축제 종합평가보고서를 발간하고 있으며 보고서에는 축제에 대한 만족도, 방문객의 소비지출, 외지관광객 비율 등에 대한 지자체의 자체 설문결과가 제시되어 있다. 매년 개최되는 전국의 축제는 계속해서 증가하는 추세이며 2020년 초에 계획된 축제도 900개 이상으로 나타났다(문화체육

* 이 논문은 2019년 대한민국 교육부와 한국연구재단의 지원을 받아 수행된 연구임 (NRF-2019S1A3A2098438)

** 경희대학교 스마트관광연구소, keunmi100@gmail.com(주저자)

*** 부산대학교 경영대학, hongth@pusan.ac.kr(교신저자)

관광부, 2020). 그러나 이 중에서 문화관광축제 로 선정되는 축제는 10% 미만으로 극소수에 불과하다. 이처럼 대부분의 축제는 평가의 사각지대에 존재하여 체계적인 평가시스템의 적용이 어려우며 이에 따른 개선이 이루어지기 어렵다. 지역관광의 활성화를 위해서는 문화관광축제로 선정되지 않은 축제에 대해서도 평가를 통한 지속적인 관심과 지원이 필요하다.

또한, 문화체육관광부가 매년 실시하고 있는 평가시스템은 선정규모의 변동성이 크며 지원 예산에 대한 편차로 인한 불확실성이 크고 투입과 축제규모에 대한 차별적인 평가를 고려하지 않아 평가체계 결과에 대한 신뢰성 및 타당성에 대한 문제가 지적되고 있다(전예빈, 2017). 이러한 문제점을 해결하기 위해 DEA(Data Envelopment Analysis)를 적용하여 문화관광축제를 보다 객관적으로 평가하고자 하였으나 대부분의 연구는 DEA를 통해 효율적인 축제와 비효율적인 축제로 나누어 원인을 파악하고 비효율적인 축제를 위한 벤치마킹 집단을 제공하며(김용묵, 조동현, 2019), 효율성 값을 토대로 전략을 제시하고 있다(전예빈, 2017). 그러나 DEA는 상대적 효율성을 평가하기 때문에 새로운 평가 대상이 추가될 때마다 DEA 모형을 다시 적용해야 하는 번거로움이 있다. 즉 문화관광축제로 선정된 축제에 대한 효율성을 판단하기 위해 DEA 모형을 매년 다시 적용해야 한다. 따라서 본 연구에서는 문화관광축제의 효율성을 평가하여 등급을 분류하고, 효율성 등급을 예측하는 모형을 제안한다. 따라서 효율성 등급 예측모형을 이용하면 효율성 평가 때마다 DEA 모형을 적용할 필요가 없으며 평가의 사각지대에 존재하고 있는 대다수의 지역축

제에 적용할 수 있다.

본 연구에서는 문화체육관광부에서 제공하는 문화관광축제로 선정된 최근 5년간의 데이터와 매년 발간하는 문화관광축제 종합평가 보고서, 관광지식정보시스템의 자료를 이용하여 문화관광축제의 효율성 등급 예측모형을 개발한다. 먼저 효율성 평가를 위해 효율성 분석의 대표적인 방법의 하나인 DEA를 적용하여 문화관광축제의 상대적 효율성을 분석한다. 축제의 규모를 반영하기 위해 BCC모형을 적용하며 이때 효율점수를 기반으로 축제를 4개의 등급으로 분류하여 기존의 효율과 비효율의 이진분류가 아닌 4등급 체계의 다분류를 통해 보다 세분화된 정보를 반영할 수 있게 한다. 문화관광축제의 효율성 등급 예측모형을 개발하는 과정에서 대표적인 기계학습 기법인 인공신경망과 Support Vector Machines(SVMs)을 적용한다. 또한, SVM의 초평면과 마진에 기반한 이진분류 한계를 극복하기 위해 다분류 알고리즘인 OAO(One-Against-One), OAR(One-Against-Rest), C&S(Crammer & Singer) 등을 적용하여 4등급 체계인 효율성 등급 예측모형을 개발한다. 본 연구에서 제안하는 효율성 등급 예측모형은 효율과 비효율의 이진분류가 아닌 다분류로 보다 세분화된 정보를 제공하며, 새로운 평가 대상이 추가되더라도 DEA 모형을 다시 적용하지 않고 축제의 효율성 등급을 예측모형을 통해 빠르게 평가가 이루어질 수 있다. 또한, 기존의 문화관광축제로 선정되지 못한 축제에도 적용하여 효율성 등급을 예측 및 평가하여 축제의 지속적인 성장을 위한 방안을 마련할 수 있으며 문화관광축제에 대한 평가를 통해 지역관광의 활성화를 위한 정책 수립에도 기여할

것으로 판단된다.

II. 이론적 배경

2.1 문화관광축제

문화관광축제사업은 문화체육관광부에서 추진하고 있으며 관광 상품성이 큰 지역축제를 대상으로 지원 및 육성하고 있다. 문화관광축제는 각 광역자치단체에서 추천받아 관광, 여행, 문화, 예술, 콘텐츠 등 다양한 분야의 전문가로 구성된 선정위원회에서 축제 프로그램의 콘텐츠, 축제 운영, 발전 가능성 등을 기준으로 선정한다(문화체육관광부, 2018). 1996년에는 문화관광축제로 8개가 선정되었으며 선정되는 축제가 점차적으로 증가하여 2013년 42개, 2014년 40개, 2015년 44개, 2016년 43개, 2017년 41개, 2018년 41개로 매년 40개 정도의 축제가 문화관광축제로 선정되고 있다.

선정된 문화관광축제는 대표축제, 최우수축제, 우수축제, 유망축제로 축제를 세분화하여 등급별로 차등 지원을 받게 된다. 문화관광축제의 질적 제고와 세계적인 축제 육성을 위해서는 객관적이고 공정한 축제의 평가체계를 구축하여야 한다. 또한, 축제의 평가에는 많은 전문가의 노력과 자원이 투입된다. 평가는 전문가 현장 모니터링, 지자체 자체 성과평가, 종합평가 등으로 이루어진다. 전문가 현장 모니터링 방식은 평가위원이 2일 이상 체류하며 각자 전문지식을 기반으로 독립적으로 판단하여 점수를 부여한다. 소비자 평가는 축제별 인지도 및 축제 만족도에 대한 온라인 설문조사로 이루어

지고 지자체 자체성과 평가는 방문객 만족도 조사와 지출액을 통한 경제효과를 분석하며 종합평가는 각 축제별 기획 목적 및 방향을 고려하여 설정한 자체 성과지표 결과를 반영한다(문화체육관광부, 2018).

문화관광축제에 대한 평가는 축제의 목적 혹은 경제적 효과의 달성에 대한 효과성 평가와 투입대비 산출에 대한 효율성 평가로 이루어질 수 있다. 축제를 평가하기 위해 개별축제에 대한 서비스품질 또는 만족도(김기현, 윤유식, 2015; 박철호, 한수정, 2014), 지역축제에 대한 경제적 파급효과(안소현 등, 2018; 김한주, 2016; 이정훈, 김성우, 2014)와 같은 효과성에 대한 연구를 중심으로 이루어지고 있으며 축제의 규모를 반영하여 투입대비 산출에 대한 차별적인 성과를 보는 효율성에 대한 연구(김용목, 조동혁, 2019; 전예빈, 2017)는 상대적으로 부족하다. 문화관광축제의 평가체제에 대해 지속적으로 지적되고 있는 예산의 낭비와 같은 문제를 해결하기 위해 투입에 대한 축제의 효율성에 대한 평가가 이루어져야 하며 본 연구에서는 DEA를 적용하여 효율성 기반으로 축제를 평가하고자 한다.

2.2 효율성 모형

효율성은 성과의 측정 및 평가를 위해 활용되고 있으며 투입 대비 산출의 비율로 정의된다. 효율성을 측정하는 방법 중 하나인 DEA(Data Envelopment Analysis)는 비교가 가능한 의사결정단위(Decision Making Units, DMU)간의 상대적 효율성을 평가하기 위한 방법이다. DEA는 비모수 방법으로 투입요소와

산출요소에 대한 제약이 없어 상대적 효율성 및 생산성을 비교하기 위한 다양한 분야에 적용되고 있다. 홍한국 등(2006)은 SI 프로젝트의 효율성 평가를 위해 DEA를 적용하였으며 DEA의 단점을 개선하기 위해 Tier 분석을 적용하여 효율성의 크기별로 의사결정단위들을 층별화하여 적용하고 기계학습을 결합한 모형을 제안하였다. 박지영과 홍태호(2009)는 국내 벤처기업의 상대적 효율성을 분석하여 다분류 효율성 등급모형을 개발하였다. 관광분야에서도 효율성을 측정하기 위해 DEA가 적용되고 있으나 효율과 비효율의 이진으로만 분류하고 있으며 관광업체의 효율성 분석을 위한 연구(송석만, 한옥상, 2015; Assaf, 2012)와 지역축제의 효율성 분석(김용목, 조동혁, 2019; 전예빈, 2017) 등에 적용되었다.

DEA 모형은 투입과 산출요소에 따라 효율성이 달라질 수 있어 변수의 선정이 중요하다. 김용목, 조동혁(2019)은 총예산, 지방예산비율을 투입요소로, 1일 평균관람객, 1인당 지출액, 외지관광비율을 산출요소로 하여 문화관광축제의 효율성을 분석하였다. 효율점수를 기반으로 효율적인 축제와 비효율적인 축제로 나누어

원인을 파악하고 비효율적인 축제를 위한 벤치마킹 집단을 제공하였다. 전예빈(2017)은 효율성을 활동 효율성과 효과 효율성으로 나누어 평가하였다. 활동 효율성을 위한 투입요소는 총예산, 개최기간, 개최횟수, 산출요소는 1인당 소비지출, 외지관광객 수, 지역주민 수, 경제효과로 하였으며 효과 효율성을 위한 최종 산출물은 경제효과로 선정하였으며 효율성에 대한 결과값을 토대로 4개의 집단으로 분류하여 각 집단에 맞는 전략을 제공하였다. 최강화(2016)는 문화관광축제의 상대적 효율성을 분석하기 위해 투입요소로 총예산을 선정하였으며 산출요소로 관광객 수와 경제효과를 이용하여 메타프론티어 분석을 수행하였다. 강인규 등(2012)의 연구에서도 문화관광축제의 효율성을 분석하기 위해 DEA를 적용하였으며 투입요소를 총예산, 지방예산비율로 하였으며, 산출요소를 1일 관람객 수, 1인당 지출액으로 하여 효율적인 축제와 비효율적인 축제를 파악하고 비효율적인 축제에 대한 벤치마킹의 대상을 단계별로 제시하기 위해 Tier 분석을 적용하였다. <표 1>은 지역축제의 효율성을 평가하기 위해 사용한 투입요소와 산출요소이다.

<표 1> 지역축제의 효율성 평가

연구자	투입요소	산출요소	DMU
김용목, 조동혁 (2019)	총예산, 지방예산비율	일평균관람객, 1인당지출액, 외지관광비율	문화관광축제
전예빈(2017)	총예산, 개최기간, 개최횟수	1인당 소비지출, 외지관광객수, 지역주민 수, 경제효과	문화관광축제
최강화(2016)	총예산	관광객수, 경제효과	지역축제
강인규 등(2012)	총예산, 지방예산비율	1일 관람객수, 1인당 지출액	문화관광축제
김재윤 등(2010)	예산	1일 평균 외부관광객 수, 외부관광객 1인당 지출액	전라남도 지역축제

일반적으로 DEA를 적용하기 위해 CCR (Charnes-Copper-Rhodes) 모형과 BCC (Banker-Charnes-Cooper) 모형이 많이 활용된다. Charnes et al.(1978)에 의해 개발된 CCR 모형은 규모의 수익성(return to scale)을 고려하지 않고 효율성을 평가하는 방법으로 규모에 대한 수익불변이라는 가정하에 효율성을 분석한다. 따라서 CCR 모형은 규모의 효율성(scale efficiency)과 순수 기술적 효율성(technical efficiency)을 구분하지 못하는 문제가 존재한다. 규모의 불변을 가정한 CCR 모형의 대안으로 개발된 BCC 모형은 Banker et al.(1984)에 의해 개발되었다. BCC 모형은 규모 효율성과 순수 기술 효율성을 측정할 수 있어 CCR 모형에서 나타난 비효율성이 기술적인 원인에 의한 것인지 규모의 원인에 의한 것인지 판단할 수 있다. BCC 모형은 CCR 모형에 불록성의 제약 조건을 추가하여 DMU를 평가하기 때문에 규모를 고려한 효율성을 평가할 수 있다. 투입 중심의 BCC 모형은 DMU의 투입 가중합에 대한 산출 가중 합의 비율을 최대화하는 선형분수계획법으로 식(1)과 같이 나타낼 수 있다.

$$\begin{aligned} \min \theta_k \\ \text{subject to } \theta_k x_k &\geq \sum_{j=1}^n x_j \lambda_j \\ \sum_{j=1}^n y_j \lambda_j &\geq y_k \\ \sum_{j=1}^n \lambda_j &= 1 \\ \lambda_j &\geq 0, j = 1, 2, \dots, n \end{aligned} \quad \text{식(1)}$$

2.3 다분류 모형

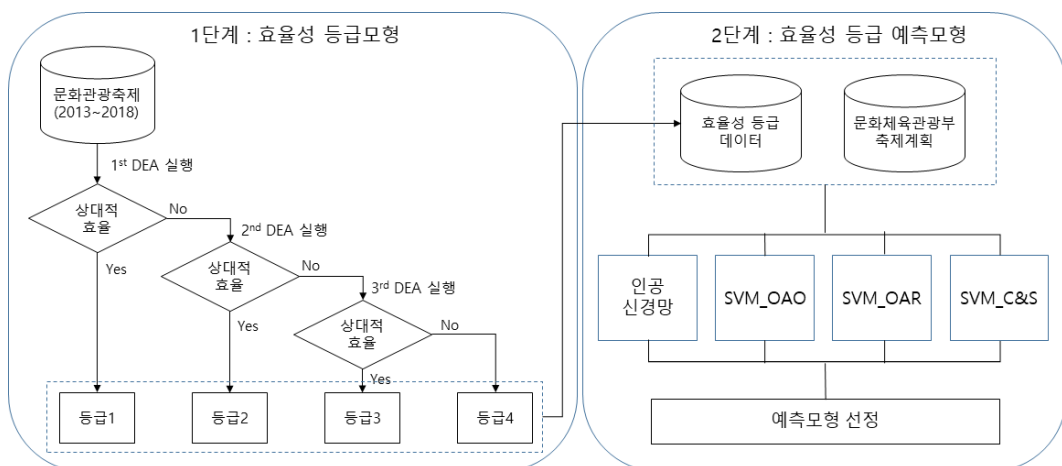
다분류 모형은 종속변수가 3개 이상의 등급으로 구분된 경우로 이진분류보다 세분화된 정보를 제공할 수 있다. 본 연구에서는 4개의 등급으로 분류된 문화관광축제의 효율성 등급을 예측하기 위해 다분류 기법을 적용하고자 한다. 축제의 효율성 등급을 다분류 문제로 설정하고 이를 해결하기 위한 기계학습법으로 인공신경망과 SVM(Support Vector Machines)을 적용하고자 한다. 인공신경망은 뇌에서 뉴런들이 상호작용하고 학습하는 것을 모방한 것으로, 분류 및 예측을 위한 다양한 분야에 적용되고 있으며 다분류 문제에도 많이 활용되고 있다(김은미, 박지영, 2019). Vapnik(1995)에 의해 제안된 SVM은 통계적 학습이론을 기반으로 하며 데이터 분류문제에 대해 일반화 능력이 높아 많이 활용되고 있다(Tay and Cao, 2001). 원하람 등(2018)의 연구와 Nguyen and Lee (2017)의 연구에서는 고객분류를 위해 SVM을 적용하였으며 다른 기계학습보다 높은 성과를 보였다. 다분류 SVM 기법은 전통적인 SVM모형을 변형 혹은 확장시켜서 사용한다. 다분류 SVM은 크게 두 가지 방법으로 구분되는데, 첫 번째는 다분류 문제를 여러 개의 이진분류 문제로 쪼개고 결합하는 방법이며, 두 번째는 한번에 모든 데이터를 고려하여 최적화 문제로 해결하는 방법이다(Hsu and Lin, 2002). 이진분류 기반의 방법 중 하나인 OAO(One-Against-One)은 k개의 집단을 두 개의 집단씩 쌍을 이루도록 하여 ${}_k C_2 = k(k-1)/2$ 개의 이진분류 모형을 생성한다. 각 분류기의 투표결과 가장 많은 표를 획득한 집단을 최종적으로 선택

하여 분류할 수 있다. OAR(One-Against-Rest)은 OAA(One-Against-All)라고도 하며 k개의 집단에 대해 하나의 집단과 나머지 집단으로 총 k개의 이진분류 모형을 구축하는 방법으로 개념적으로는 가장 단순한 다분류 기법이다 (Ahn and Kim, 2009). 모든 데이터를 한 번에 고려하는 C&S(Crammer & Singer)방법은 분류기를 찾는 과정이 매우 복잡하여 시간이 오래 걸리며 많은 컴퓨터 자원을 필요로 한다는 단점이 있다(Hsu and Lin, 2002; Lorena and Carvalho, 2008). 본 연구에서는 인공신경망과 다분류 SVM 기법을 적용하여 문화관광축제의 효율성 등급을 효율과 비효율의 이진분류가 아닌 4개 등급을 분류하는 다분류 기법을 적용하고자 한다.

Ⅲ. 연구 프레임워크

본 연구에서는 문화관광축제의 효율성을 분석하고 효율성 등급 예측모형을 구축하고자 한

다. 먼저 최근 5년간 문화관광축제로 선정된 축제의 효율성을 평가하기 위해 DEA를 적용한다. 축제의 규모를 반영하기 위해 BCC 모형을 적용하여 상대적인 효율성을 평가하며 DEA는 단순 효율성 평가가 아닌 총화기법의 DEA를 통해 상대적 효율점수를 기반으로 문화관광축제를 총별화하여 적용한다. 의사결정단위의 총별화 방법은 Thanassoulis(1995)에 의해 제안되었으며, 홍한국 등(2006)과 Park and Hong (2009)의 연구에서는 DEA에 의한 상대적인 효율점수를 기반으로 효율과 비효율을 분류하고 비효율적인 의사결정단위를 대상으로 다시 DEA를 적용하여 총별화하였다. 본 연구에서는 총화기법의 DEA를 적용하여 문화관광축제를 4개의 등급으로 분류하고 효율성 등급을 예측하기 위한 다분류 모형을 구축한다. 효율성 등급의 다분류 예측을 위해 인공신경망과 SVM의 다분류 기법인 OAO(One-Against-One), OAR(One-Against-Rest), C&S(Crammer & Singer)를 적용하여 SVM_OAO, SVM_OAR, SVM_C&S 모형을 구축하며 예측성과를 비교



<그림 1> 연구 프레임워크

한다. DEA를 적용한 효율성 평가는 상대적인 효율성을 평가하기 때문에 새로운 축제가 문화관광축제로 선정되면 DEA를 다시 실행해야 하는 번거로움이 있다. 또한, 매년 개최되는 축제 중 문화관광축제로 선정되는 비율은 10% 미만으로 문화관광축제로 선정되지 못하면 평가의 대상이 되지 않는다. 반면에 본 연구에서 제안하는 효율성 등급 예측모형은 새로운 축제가 문화관광축제로 선정되더라도 효율성 평가를 재실행하지 않아도 되며 문화관광축제로 선정되지 않은 축제의 효율성 평가를 위해서도 적용이 가능하다. 본 연구의 연구 프레임워크는 <그림 1>과 같다.

IV. 분석방법 및 결과

4.1 데이터

본 연구에서는 문화체육관광부에서 제공하는 문화관광축제 종합보고서와 지역축제자료, 관광지식정보시스템의 통계자료를 활용하였다. 문화관광축제 종합보고서는 각 축제의 방문객을 대상으로 만족도 조사의 분석결과와 소비지출, 외지관광객 비율 등의 정보를 제공하고 있다. 관광지식정보시스템에서는 내국인 및 외국인 방문객 수, 축제 개최 후 발생한 경제효과에

대한 데이터를 제공하고 있으며 지역축제 자료에서는 개최계획에 대한 정보를 제공하며 예산 및 개최기간, 전년도의 방문객 수 등 축제에 대한 전반적인 정보를 제공한다.

2013년부터 2018년까지 문화관광축제로 선정된 축제는 251개이며 통계적으로 큰 오차를 보이는 축제와 만족도 조사를 실시하지 않은 축제 등 충분한 데이터를 보유하지 않은 데이터를 제외하고 총 216개의 데이터를 분석에 사용하였다.

4.2 문화관광축제의 효율성 분석

효율성 분석에서 가장 많이 사용되는 DEA를 적용하여 문화관광축제의 효율성을 평가하였다. DEA를 실행하기 위해서는 투입변수와 산출변수의 선정이 중요하며 본 연구에서는 투입변수를 총예산, 축제 개최기간으로 설정하였으며, 산출변수는 1일 평균 방문자 수, 전체 만족도로 선정하였다. 총방문자 수는 축제 개최기간에 영향을 받을 것으로 판단하여 1일 평균 방문자 수를 적용하였으며 DEA의 투입변수와 산출변수의 기초통계량은 <표 2>와 같다. 투입변수와 산출변수의 편차가 크게 나타나며 이는 축제의 상대적인 규모로 인한 것으로 추정된다. 따라서 축제의 규모를 반영한 상대적 효율성을 평가하기 위해 BCC모형에 대한 DEA를 실행

<표 2> DEA의 투입·산출변수

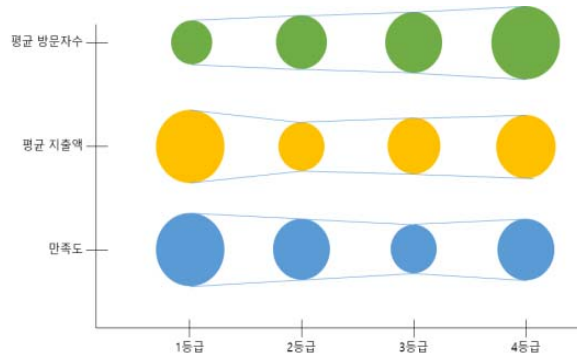
구분		평균	표준편차
투입변수	총예산(백만원)	1,034.7	528.7
	개최기간(일)	5.9	3.6
산출변수	1일 평균 방문자 수(명)	76,709.3	69,190.8
	전체 만족도	5.2	0.4

하였다.

DEA를 적용하기 위해 분석도구 R을 적용하였다. DEA실행 결과 216개의 문화관광축제 중 43개의 축제가 BCC모형의 효율점수가 1인 것으로 나타났다. 층화기법의 DEA를 적용하여 첫 번째 DEA에 의해 상대적으로 효율적인 43개의 축제에 효율성 등급 1을 부여하였다. 효율성 등급 1이 부여된 43개의 축제를 제외하고 두 번째 DEA를 실행한 결과 BCC모형의 효율점수가 1인 축제는 62개이며 효율성 등급 2를 부여하였다. 동일한 방법으로 효율성 등급이 1인 축제와 2인 축제를 제외하고 DEA를 실행하여 BCC모형의 효율점수가 1인 축제는 3등급으로 나머지는 4등급을 부여하여 층별화 방법에 의해 문화관광축제의 효율성 점수를 기반으로 4개의 등급으로 분류하였으며 분류된 축제는 <표 3>과 같다.

각 등급에 대한 방문자의 특성을 살펴보기 위해 평균 방문자 수, 평균 지출액, 만족도를 비교해본 결과 <그림 2>와 같다. 효율성 등급에 따라 만족도에서는 큰 차이를 보이지 않으며 평균

방문자 수는 4등급에서 평균 지출액은 1등급에서 가장 높게 나타났다. 1등급은 가장 효율적인 축제로 평균 방문자 수는 적지만 평균 지출액과 만족도는 높게 나타나 효율성 등급이 유지될 수 있도록 지원을 해야 할 것으로 보인다. 2등급은 평균 지출액이 가장 적은 것으로 나타났으며 3등급은 만족도가 가장 낮은 특성을 보이며 이에 대한 원인을 찾아 대책을 마련해야 할 것이다.



<그림 2> 효율성 등급에 따른 방문자의 특성

<표 3> BCC모형을 적용하여 4개의 등급으로 분류

효율성 등급	효율 축제 수	설명	평균 방문자 수 (명)	평균 지출액 (원)	만족도
1등급	43	첫 번째 DEA 실행 결과 BCC 모형의 효율점수가 1인 축제	267,164.37	62,860.72	5.32
2등급	62	첫 번째 DEA 실행 결과 효율점수가 1인 축제 (1등급)를 제외하고 두 번째 DEA를 실행한 결과 효율점수가 1인 축제	372,003.53	53,229.92	5.13
3등급	48	두 번째 DEA 실행 결과 효율점수가 1인 축제 (2등급)를 제외하고 세 번째 DEA를 실행한 결과 효율점수가 1인 축제	460,713.45	56,594.98	5.06
4등급	64	세 번째 DEA 실행 결과 효율점수가 1인 축제 (3등급)를 제외한 나머지 축제	610,020.70	59,668.73	5.13

4.3 기계학습을 이용한 축제 효율성 등급 예측

본 연구에서는 효율성 등급을 예측하기 위해 총화기법의 DEA를 적용하여 문화관광축제를 4등급으로 분류하고 DEA 기반의 효율성 등급 예측모형을 개발하였다. 예측모형은 새로운 축제에 대해서도 효율성 등급을 예측할 수 있으며 문화관광축제로 선정되지 않은 축제에도 적용할 수 있다. 문화체육관광부에서 매년 발표하는 지역축제 개최계획에는 국비, 시·도비, 구·군비, 기타 등으로 이루어진 총예산이 제공된다. 축제 개최 전 파악할 수 있는 올해의 총예산과 예산에서 시·도비와 구·군비로 이루어진 지방자치단체의 지방예산비율, 축제개최횟수, 개최일수, 전년도의 방문객 수의 정보를 입력변수로 하여 축제의 효율성 등급을 예측하고자 한다. 예측모형을 위한 입력변수의 기초통계량은 <표 4>와 같다.

예측모형을 구축하기 위해 각 등급별로 학습용 데이터와 검증용 데이터를 4:1로 분류하여 학습용 데이터로 모형을 구축하고 선택하였으며 검증용 데이터로 모형을 평가하였다. 또한, 표본추출에 따른 예측모형의 신뢰성을 위해 5-fold cross validation을 적용하고 python으로 분석하였다. 인공지능망 모형에서는 은닉노드의 수를 1부터 15까지 변화시켜가며 성과가 가장 좋은 모형을 선정하였다. SVM 기반의 다분류 모형인 SVM_OAO, SVM_OAR, SVM_C&S를 적용하였으며 RBF 커널함수를 적용하고 파라미터인 C와 gamma 값을 변화시켜가며 성과가 좋은 모형을 선정하였다. 예측모형의 성과는 <표 5>와 같다. SVM_OAO 모형의 평균 성과가 85.70%로 가장 높게 나타났으며 SVM_OAR 모형은 73.19%로 가장 낮은 성과를 보였다. SVM_OAR 모형은 하나의 등급과 나머지 등급으로 모형을 구축하게 되어 불균형 데이터로 인해 낮은 성과를 보인 것으로 판단된다.

<표 4> 예측모형의 입력변수

변수명	평균	표준편차	최댓값	최솟값
예산(백만원)	1,042.20	522.62	3,009	7.4
지방예산비율	0.80	0.14	1	0
개최횟수(회)	20.74	12.49	60	4
개최기간(일)	5.94	3.63	23	3
전년도의 총방문객 수(천명)	472.62	445.92	2,800	51.05

<표 5> 예측모형 성과

	인공신경망		SVM_OAO		SVM_OAR		SVM_C&S	
	학습용	검증용	학습용	검증용	학습용	검증용	학습용	검증용
Set1	84.57	82.93	90.29	90.24	78.29	70.73	90.86	90.24
Set2	81.29	84.44	89.47	88.89	76.61	73.33	84.80	88.89
Set3	81.29	66.67	91.23	80.00	76.61	66.67	84.80	75.56
Set4	82.08	88.37	91.33	86.05	76.88	81.40	89.02	88.37
Set5	85.63	78.57	90.80	83.33	80.46	73.81	90.23	83.33
평균	82.57	80.20	90.62	85.70	77.77	73.19	87.94	85.28

V. 결과 및 향후 연구과제

본 연구에서는 문화관광축제의 효율성을 평가하고 이를 기반으로 효율성 등급을 예측하는 모형을 구축하였다. 효율성을 평가하기 위해 가장 많이 사용되는 DEA를 적용하였으며 축제의 규모를 반영하기 위해 BCC 모형의 효율성을 적용하였다. 효율성 등급을 분류하기 위해 단순 DEA가 아닌 층화기법의 DEA를 적용하였으며 4개의 등급으로 분류하였다. DEA에서는 투입 요소와 산출요소가 중요한 요인으로 본 연구에서는 총예산, 개최기간을 투입요소로 사용하고 1일 방문자 수, 만족도를 산출변수로 사용하여 정량적인 데이터뿐 아니라 정성적인 데이터도 반영하여 효율점수를 계산하였다. 각 등급별 평균 방문자 수, 평균 지출액, 만족도에 대한 요인을 비교해 본 결과 등급별로 만족도에는 큰 차이를 보이지 않았으나 평균 방문자의 수는 1등급에서 가장 적고 4등급에서 가장 많으며 1등급에서 평균 방문자 수는 적지만 평균 지출액과 만족도는 가장 높게 나타났으며 평균 지출액은 2등급에서 가장 낮으며 만족도는 3등급에서 가장 낮게 나타났다.

효율점수에 의해 4개의 등급으로 분류된 문화관광축제에 대해 다분류 모형을 적용하여 효율성 등급을 예측하였다. 다분류를 위해 인공신경망과 SVM 기반의 SVM_OAO, SVM_OAR, SVM_C&S 모형을 적용하였으며 매년 초에 문화관광체육부에서 발표하는 축제 개최계획의 정보인 총예산, 지방예산비율, 축제 개최횟수, 축제 개최기간, 전년도 방문객 수를 입력변수로 사용하였다. SVM_OAO 모형에서 우수한 예측 성과를 보였으며 SVM_OAR 모형에서는 다른

모형에 비해 낮은 성과를 보였으며 이는 학습 데이터에서 효율과 비효율 데이터 분포의 불균형도가 높았기 때문인 것으로 판단된다.

본 연구의 시사점은 다음과 같다. 첫째, 본 연구는 효율성을 효율과 비효율의 이진분류가 아닌 층화기법의 DEA를 적용하여 다분류를 적용하였다. 이는 보다 세분화된 정보를 제공하며 각 세분화된 등급별 특성에 맞는 전략을 수립할 수 있다. 효율성 등급이 방문객이 가장 적게 나타나는 1등급이라면 축제에 대한 홍보를 통해 방문객을 유인할 수 있는 전략을 수립할 수 있다. 둘째, 효율성 등급 예측모형을 구축하여 DEA를 매년 재실행할 필요없이 문화관광축제의 효율성 등급을 예측할 수 있다. 기존의 DEA는 새로운 축제가 추가되거나 매년 선정되는 문화관광축제에 대해 DEA를 다시 실행해야 하지만 예측모형을 통해 재실행에 대한 번거로움을 줄일 수 있다. 셋째, 문화관광축제로 선정되지 않은 지역축제에도 적용할 수 있어 평가의 사각지대에 놓여있는 지역축제에도 쉽게 적용할 수 있다. 문화관광축제로 선정되지 않더라도 효율성 등급 예측모형을 적용하여 자체적인 평가가 가능하며 문제점을 미리 파악하고 개선하여 전략을 수립할 수 있으며 지역사회의 활성화에도 기여할 것으로 판단된다.

본 연구의 한계점 및 향후 연구방향은 다음과 같다. 첫째, DEA 모형에서는 투입요소와 산출요소에 따라 효율성이 달라질 수 있어 향후 연구에서는 투입과 산출요소의 선정을 위해 보다 명확한 근거가 필요할 것으로 보인다. 둘째, 개최 전 축제에 대해 파악할 수 있는 정보가 적어 뉴스정보와 같은 축제와 관련된 뉴스정보와 축제를 방문했던 고객리뷰 정보 등 다양한

변수를 적용할 수 있다면 보다 신뢰성 있는 성과를 얻을 수 있을 것으로 판단되며 향후 연구에서는 문화관광축제에 대한 텍스트 데이터를 예측모형에 활용하며 보다 다양한 모형을 구축해 볼 수 있을 것이다.

참고문헌

- 강인규, 김재윤, 이수현, “DEA Window 모형과 Tier 분석을 이용한 축제의 효율성 분석”, 기업경영연구, 제19권, 제1호, 2012, pp. 63-82.
- 김기현, 윤유식, “문화관광축제 평가속성과 방문객 만족, 충성도, 이미지개선 구조적 관계분석, 관광레저연구, 제27권, 제5호, 2015, pp. 185-200.
- 김용목, 조동혁, “국내 문화관광축제의 효율성 분석”, 관광레저연구, 제31권, 제6호, 2019, pp. 167-187.
- 김은미, 박지영, “온라인 P2P 환경에서 딥러닝을 적용한 다분류 기반 개인신용등급 예측모형, 인터넷전자상거래연구, 제19권, 제4호, 2019, pp. 43-57.
- 김한주, “지역산업연관모형을 이용한 광역시도별 관광산업의 경제적 파급효과 추이에 관한 연구”, 관광레저연구, 제28권, 제5호, 2016, pp. 61-81.
- 문화체육관광부, “2018 문화관광축제 종합평가 보고서”.
- 문화체육관광부, “2020 문화관광축제 종합평가 보고서”.
- 박철호, 한수정, “문화관광축제 평가속성과 방문객의 만족, 행동의도에 관한 연구: 충주세계무술축제를 중심으로”, 관광경영학회, 제18권, 제1호, 2014, pp. 81-100.
- 송석만, 한옥상, “서비스무역으로서의 외국인 이용 관광호텔의 효율성 분석”, 무역학회지, 제40권, 제4호, 2015, pp. 217-238.
- 안소현, 이충기, 이새미, 김현래, “지역산업연관모형을 이용한 서울빛초롱축제의 경제적 파급효과 분석”, 관광연구저널, 제32권, 제5호, 2018, pp. 51-63.
- 원하람, 김무전, 안현철, “온라인 무료 샘플 판촉의 효과적 활용을 위한 기계학습 기반 고객분류예측 모형”, 정보시스템연구, 제27권, 제3호, 2018, pp. 63-80.
- 이정훈, 김성우, “지역산업연관모형을 활용한 지역축제의 경제적 파급효과 분석_의성마늘국제연날리기대회를 중심으로”, 한국지방자치연구, 제16권, 제2호, 2014, pp. 351-374.
- 전예빈, “DEA를 이용한 문화관광축제의 효율성 평가”, 석사학위논문, 한양대학교, 2017.
- 최강화, “메타프론티어 분석을 이용한 지역축제의 효율성 비교”, 관광연구, 제31권, 제6호, 2016, pp. 27-46.
- 홍한국, 김종원, 서보라, “SI 프로젝트의 효율성 평가를 위해 자료포괄분석과 기계학습을 결합한 하이브리드 분석”, 한국산업정보학회 학술대회논문집, 2006, pp. 77-88.
- Ahn, H. and Kim, K. J., “Corporate Bond Rating Using Various Multiclass Support Vector Machines,” *Asia Pacific*

- Journal of Information Systems*, Vol. 19, No. 2, 2009, pp. 157-178.
- Assaf, A. George., "Benchmarking the Asia Pacific tourism industry: A Bayesian combination of DEA and stochastic frontier," *Tourism Management*, Vol. 33, No. 5, 2012, pp. 1122-1127.
- Banker, R.D., Charnes A., and Cooper, W.W., "Some Models for Estimating Technical and Scale Inefficiencies in Data Envelopment Analysis," *Management Science*, Vol. 30, No. 9, 1984, pp. 1078-1092.
- Charnes, W.W., Cooper, and Rhodes, E., "Measuring the Efficiency of Decision Making Units," *European Journal of Operation Research*, Vol. 2. No. 6, 1978, pp. 429-444.
- Hsu, C. W. and Lin, C. J., "A Comparison of Methods for Multiclass Support Vector Machines," *IEEE Transactions on Neural Networks*, Vol. 13, No. 2, 2002, pp. 415-425.
- Lorena, A. C. and de Carvalho, A. C. P. L. F., "Evolutionary tuning of SVM parameter values in multiclass problems," *Neurocomputing*, Vol. 71, 2008, pp. 3326-3334.
- Nguyen, P. and Lee, Y., "An Application of Support Vector Machines to Customer Loyalty Classification of Korean Retailing Company Using R Language," *정보시스템연구*, 제26권, 제4호, 2017, pp. 17-37.
- Park, J. and Hong, T., "The Prediction of DEA based Efficiency Rating for Venture Business Using Multi-class SVM," *Asia Pacific Journal of Information Systems*, Vol. 19, No. 2, 2009, pp. 139-155.
- Tay, F. E. H. and Cao, L. J., "Application of support vector machines in financial time series forecasting," *Omega*, Vol. 29, No. 4, 2001, pp. 309-317.
- Thanassoulis, E., "Assessing polic forces in Engliand and Wales using data envelopment analysis," *European Journal of Operational Research*, Vol. 87, No. 3, 1995, pp. 641-657.
- Vapnik, V., *The nature of statistical learning theory*, New York, Springer-Verlag, 1995.

김 은 미 (Kim, Eun-Mi)



부산대학교 경영학과에서 석사학위와 박사학위를 취득하였다. 현재 경희대학교 스마트관광연구소에서 학술 연구교수로 재직하고 있다. 주요 관심분야는 비즈니스 애널리틱스, 딥러닝, 소셜미디어, CRM 등이다.

홍 태 호 (Hong, Tae-Ho)



부산대학교 경영대학 교수로 재직하고 있다. KAIST에서 경영정보시스템을 전공하여 공학석사와 공학박사를 취득하였다. 주요관심분야는 비즈니스 애널리틱스, 딥러닝, 오피니언 마이닝, CRM 등이다.

<Abstract>

The Efficiency Rating Prediction for Cultural Tourism Festival Based of DEA

Kim, Eun-Mi · Hong, Tae-Ho

Purpose

This study proposed an approach for predicting the efficiency rating of the cultural tourism festivals using DEA and machine learning techniques. The cultural tourism festivals are selected for the best festivals through peer reviews by tourism experts. However, only 10% of the festivals which are held in a year could be evaluated in the view of effectiveness without considering the efficiency of festivals.

Design/methodology/approach

Efficiency scores were derived from the results of DEA for the prediction of efficiency ratings. This study utilized BCC models to reflect the size effect of festivals and classified the festivals into four ratings according the efficiency scores. Multi-classification method were considered to build the prediction of four ratings for the festivals in this study. We utilized neural networks and SVMs with OAO(one-against-one), OAR(one-against-rest), C&S(crammer & singer) with Korea festival data from 2013 to 2018.

Findings

The number of total visitors in low efficient rating of DEA is more larger than the number of total visitors in high efficient ratings although the total expenditure of visitors is the highest in the most efficient rating when we analyzed the results of DEA for the characteristics of four ratings. SVM with OAO model showed the most superior performance in accuracy as SVM with OAR model was not trained well because of the imbalanced distribution between efficient rating and the other ratings. Our approach could predict the efficiency of festivals which were not included in the review process of culture tourism festivals without rebuilding DEA models each time. This enables us to manage the festivals efficiently with the proposed machine learning models.

Keyword: DEA, Cultural Tourism Festival, Multi-class Prediction, Machine Learning

* 이 논문은 2020년 6월 1일 접수, 2020년 7월 3일 1차 심사, 2020년 9월 23일 게재 확정되었습니다.