

# 빅데이터 분석을 통한 통합문화이용권 관리 개선 방안 연구

## A Study on Ways to Improve Management of the Integrated Culture Voucher Program using Big Data Analysis

박규동<sup>1</sup>, 주성돈<sup>2</sup>, 황민철<sup>3\*</sup>

Kyu-Dong Park<sup>1</sup>, Sung-Don Ju<sup>2</sup>, Min-Chul Hwang<sup>3\*</sup>

### 요 약

본 연구는 통합문화이용권 관리방식에 대한 분석과 진단을 토대로 좀 더 체계적인 가맹점 관리방안을 모색하고자 한다. 이를 위해, 통합문화이용권 가맹점 빅데이터를 사용하여 가맹점 관리자의 의사결정 패턴을 탐색하기 위한 의사결정나무모형을 분석에 활용하였다. 분석 결과 가맹점 승인/해지 의사결정에 영향을 미치는 주요 변수는 첫째, 대분류나 중분류 기준보다 소분류 기준이 상대적으로 의사결정에 더 많은 영향을 미치고 있었다. 둘째, 결제 승인 건수와 승인 금액이 의사결정에 영향을 미치는 것으로 나타났다. 셋째, 온라인 가맹점에 관한 의사결정은 오프라인 가맹점 보다 상대적으로 단순한 패턴을 보인다. 넷째, 관광(대분류)-숙박(중분류)-숙박(소분류)으로 분류되는 가맹점에 대한 의사결정이 가장 복잡한 것으로 나타났다. 이러한 결과를 바탕으로, 본 연구는 통합문화이용권 관리를 위한 두 가지의 실질적인 방안을 제시하였다. 첫째, 가맹점 관리규정을 개편하여 관리상의 효율성을 개선하는 것이다. 둘째, 가맹점 분류체계를 좀 더 세분화 하는 것이다.

핵심어 : 통합문화이용권, 빅데이터, 의사결정나무분석, 가맹점 관리

### Abstract

This study aims to find a more systematic franchise management plan based on the analysis and diagnosis of the integrated culture voucher management. To this end, a decision tree model is used for analysis to explore the decision-making patterns of affiliate managers using the Integrated Cultural Voucher Program (ICVP) big data. In this study, decision tree decision analysis was performed to explore users' usage patterns based on big data derived from the affiliate management of ICVP. The purpose of this study is to find ICVP's management plan reflecting users' usage and decision-making patterns. As a result of the analysis, the important variables influencing the decision-making for approval/release of affiliates are as follows. First, the subcategory criteria had a greater influence on decision making than the

1 Department of Public Administration, Kwangwoon University, Seoul, Korea [Assistant Professor]  
e-mail: gdpark29@kw.ac.kr

2 Department of Public Administration, Myongji College, Seoul, Korea [Assistant Professor]  
e-mail: hyripa@mjc.ac.kr

3 School of Social Welfare, Sunchon National University, Suncheon, Korea [Associate Professor]  
e-mail: mchwang@scnu.ac.kr (Corresponding author)

\*한국문화예술위원회의 정책용역 보고서 내용을 수정 보완하였음(과제번호: ARKO 수시연구 2021-04).

\*이 논문은 순천대학교 교연비 사업에 의하여 연구되었음.

Received(May 11, 2022), Review Result(1st: May 27, 2022), Accepted(June 10, 2022), Published(June 30, 2022)



© 2022 The Authors. Published by NCSS.  
This is an open access article licensed under the Creative Commons Attribution-NonCommercial 4.0 International License.  
To view a copy of this license, visit <http://creativecommons.org/licenses/by-nc/4.0/>.

high/intermediate criteria. Second, the number of payment approvals and the amount of approval affect decision-making. Third, decision-making for online affiliates shows a relatively simple pattern compared to offline affiliates. Fourth, the decision-making process of affiliates classified into tourism (major classification level)-accommodations 1 (medium classification level)-accommodations 2 (minor class) was the most complex. Based on these results, this study suggests two practical ways to manage ICVP. First, the management efficiency is improved by reorganizing the affiliate management regulations. The second is to further subdivide the affiliate classification system.

Keyword : Integrated Cultural Voucher Program, Big Data, Decision Tree Analysis, Affiliates' Management

## 1. 서론

### 1.1 연구배경 및 목적

정보기술의 발전은 새로운 인터넷 서비스와 산업의 발전을 가속화시키고 있으며, 이로 인하여 과거 경험하지 못한 많은 정보가 생산되어지고 있다 [1]. 최근에는 빠르고, 다양한 부문에서 그리고 많은 양의 정보를 포함한 데이터를 빅데이터라 간주하고 있으며 [2], 다양한 영역에서 이러한 빅데이터를 처리, 분석, 관리에 대한 필요성이 증대되고 있다 [3].

문화 영역에서도 빅데이터를 어떻게 처리하고 분석하고 관리해야 하는지에 대한 이슈가 부각되고 있는데 주로 소비와 공급의 측면에서 논의가 이루어지고 있다 [4]. 문화의 범주가 방대하여 단일한 서비스 생산이 불가능하다는 문제점으로 인하여 그동안 다양한 소비자의 욕구와 트렌드에 대한 조사의 어려움이 존재하였다. 하지만, 최근 빅데이터 분석을 통해 문화 영역의 포괄적이고 복합적인 소비자의 욕구와 트렌드 파악 및 예측 가능성이 증가하고 있다. 이와 더불어, 이러한 빅데이터 기반의 소비 트렌드와 욕구를 반영한 공급 측면의 마케팅과 기획을 모색하여 소비자의 개인 특성에 적합한 또는 다양하고 복합적인 욕구와 트렌드를 포함한 서비스 제공 가능성이 증대되고 있어, 문화 영역에서 빅데이터 이용에 대한 관심은 확대되고 있다.

최근 문화 영역에서 빅데이터 분석을 통한 활용 방안을 제시하는 연구들이 진행되고 있으나, 문화의 특정 부문들에 한정된 연구가 주로 이루어지고 있다. 예를 들어, 공연 [5][6], 미술 [7][8], 영상 [9][10], 예술교육 [11][12], 도서 [13][14], 대중교통 [15][16], 관광지 [17][18], 숙박 [19][20], 스포츠 관람 [21][22], 체육용품 [23][24], 레저 [25][26]의 각 부문만을 한정하여 빅데이터 분석을 수행하고 있다.

지금까지 수행된 기존 연구들은 문화의 특정 범주에 국한된 빅데이터 분석만 이루어지는 한계가 있기 때문에, 본 연구는 다양한 문화 범주를 포괄할 수 있는 통합문화이용권에서 파생된 빅데이터에 대한 분석을 실시하려 한다. 통합문화이용권은 여행, 체육, 공연, 전시, 도서 등과 같은 다양한 문화 범주뿐만 아니라, 이용자의 구매정보, 카드사용 정보와 가맹점 정보가 포함된 서비스 공급자의 정보까지 생산되고 있어 빅데이터 분석에 적절하다고 알려져 있다 [4]. 따라서 본 연구의

목적은 통합문화이용권에서 파생된 빅데이터 분석을 통해 다양하고 복합적인 소비자 욕구와 사용 패턴을 발견하고, 이러한 욕구와 사용패턴을 포함한 실효성 있는 통합문화이용권 관리 방안을 모색하는 것이다.

## 1.2 통합문화이용권 개요

2022년 통합문화이용권 사업지침을 살펴보면 [27], 통합문화이용권은 저소득층의 문화, 여행, 체육 활동을 증진하여 이들의 문화 격차 완화와 삶의 질을 향상하기 위한 바우처 사업이다. 2005년 문화바우처 시범사업을 추진하여, 2006년 전국으로 확대되었고, 2013년 문화, 여행, 스포츠 관람 이용권이 통합하여 지금까지 운영되고 있다. 문화예술진흥법 15조 3, 4를 기반으로 사업의 대상자는 6세 이상의 기초생활수급자와 차상위계층이다. 이용자는 문화누리카드를 발급받아 연간 1인당 10만 원 상당의 금액을 충전 받으며, 2022년 기준 263만 명이 문화누리카드를 발급받을 수 있는 대상자이다. 이러한 바우처의 재원은 문예진흥기금과 지방자치단체의 지방비로부터 조달되고 있다. 사업추진체계는 문화체육관광부가 정책 방향과 기본계획을 수립하고 문화체육관광부 산하의 한국문화예술위원회가 사업총괄과 가맹점 관리를 하고 있다. 문화체육관광부는 각 지방자치단체에 사업계획을 통보하면 지방자치단체는 카드발급과 지역별 사업을 총괄 운영하며, 한국문화예술위원회가 문예진흥기금을 받는다. 이와 더불어, 17개 광역단체에 지역 주관처가 설립되어 사업비 집행, 관리, 지역관리가 이루어지고 있다. 문화누리카드를 이용할 수 있는 곳을 가맹점으로 관리하고 있으며, 온·오프라인 상에서 이용자가 이용할 수 있는 가맹점이 혼재되어 있다. 또한, 2022년 통합문화이용권 사업지침에 의하면 [27], 가맹점들을 [표 1]과 같이 문화, 체육, 관광의 대분류에서 중분류, 소분류로 나누어 허용된 업종과 품목에서만 카드를 이용할 수 있도록 엄격하게 관리하고 있다.

## 2. 본론

### 2.1 통합문화이용권 가맹점 분류 분석의 연구방법

본 절에서는 통합문화이용권 가맹점 등록 및 해지 기록을 분석함으로써 가맹점 관리데이터에 내재한 패턴을 체계적으로 분류하고자 한다. 통계적 기계학습(statistical machine learning)을 활용하여 분류 모형의 훈련데이터(training data)로부터 함수를 유추하고자 한다.

본 절에서는 분석자가 과거의 데이터를 활용하여 조건과 결과에 대한 매칭을 기계에게 일일이 알려주는 방식으로 학습을 시킨 후, 기계가 학습 경험을 토대로 특정 상황에서의 결괏값을 추론한 결과를 제시하는 지도학습(supervised learning)을 진행하였다 [28]. 이러한 기계학습 모형을 활용하여 분석에서는 가맹점 등록/해지 데이터를 활용하여 모형을 학습한 후, 이를 새로운 가맹점 대상에

모형을 적용하였다. 모형을 설정하기 위해 우선 기존의 개별 가맹점 등록정보를 활용하여 훈련데이터와 검증데이터(test data)로 구분하여 훈련데이터를 기계에 정답으로 학습을 시킨 후, 기계가 주어진 검증데이터를 정해진 카테고리(라벨)에 따라 분류시키는 방식을 취하였다.

[표 1] 통합문화이용권 가맹점 분류

[Table 1] Classification of Affiliates in the Integrated Culture Voucher Program

대분류	중분류 (장르)	소분류	업종 및 품목
문화	도서	도서	서점, 중고서점, 도서 대여점, 온라인서점, 만화 콘텐츠사이트, 전자책구독사이트
	음악	음악	음반 판매점, 음원 콘텐츠사이트, 악기(악기 소매점, 악기 부속품)
	영상	영화	영화관, 영상 콘텐츠 서비스(OTT 서비스 등), 영화제
		TV	케이블TV, 위성방송
	공연	공연	공연장, 극단, 예술단, 공연기획사, 아트홀, 공연 축제
	미술	전시	미술관, 박물관, 화랑, 비엔날레
		공예	공예품점, 화방(문구), 표구사
		사진관	사진관, 온라인 사진인화업체
	문화체험	문화체험	문화센터, 공방(도예 등) 및 문화예술 체험공간, 문화재 체험, 한복점, 한복대여점, VR 체험관, 방 탈출 체험, 온라인취미 클래스, 지자체 문화체험 공공서비스사이트 등
관광	교통수단	직업체험	직업체험(유·청소년에 한함)
		철도	철도(KTX, SRT, 무궁화호 등)
		시외/고속버스	시외버스, 고속버스, 공항리무진(시외 운행)
		국내 항공	항공사
		여객선	여객선
	여행사	렌터카	렌터카
		여행사	여행사
관광	관광지	관광명소	국립공원, 사적지, 시티투어, 케이블카, 모노레일, 기념관, 과학관, 천문대, 동굴, 영화(드라마) 촬영장, 산업 관광지 등
		휴양림/캠핑장	휴양림, 캠핑장, 야영장
		동·식물원	동물원, 식물원, 정원 등 동물·식물 주제 관광지
		온천	온천(온천법 허가업소)
		체험관광	지역문화축제(불임 참조), 지역문화체험(농산어촌체험 등), 템플스테이, 레일바이크, 짚라인, 생태체험 등 체험형 관광지
	숙박	테마파크	놀이공원, 워터파크, 아쿠아리움, 실내테마파크(대형), 민속촌 등
		숙박	호텔, 리조트, 콘도, 모텔, 게스트하우스, 민박, 연수원, 수련원 등
체육	스포츠 관람	스포츠 관람	4대 프로스포츠(야구, 축구, 농구, 배구) 입장권, 국내 개최 국제스포츠 경기 입장권
	체육용품	체육용품	체육사 및 체육용품점, 구단 공식 응원용품점, 자전거 판매점, 공유 자전거·전동킥보드
	체육시설	체육시설	(종목) 수영, 헬스, 볼링, 요가, 에어로빅, 필라테스, 복싱, 탁구, 사격, 롤러스케이트, 승마, 스케이트, 스키, 태권도, 합기도, 스포츠댄스, 방 송댄스 / 체육시설이용 예약플랫폼, 스크린 체육시설, 레저스포츠 등

본 연구에서는 전체 가맹점 데이터 중 30%를 훈련 데이터로 구분하여 지도학습을 시킨 후, 이

를 토대로 얻은 기계학습모형을 나머지 70% 검증데이터를 활용하여 모형의 성능을 평가하였다. 분석을 위해 문화예술위원회의 2020년, 2021년 가맹점 관리데이터를 사용하였다. 가맹점 관리데이터는 2020년과 2021년 각각 16,328개, 15,702개 가맹점 등록정보, 승인 및 해지 이력, 결제이력 등을 포함하고 있다.

이러한 지도학습 모형설정을 통해 통합문화이용권 운영주체(문화예술위원회 문화누리부)가 어떤 사업자를 통합문화이용권 가맹점으로 승인할 것인지, 승인 안 할 것인지 또는 기존의 가맹점을 해지할 것인지를 체계적으로 분류하고 설명하고자 한다. 이를 위해 기존 가맹점 등록 및 관리정보에 존재하는 승인 및 해지 이력을 이진 분류(binary classification) 문제로 구성하여 각 의사결정 범주에 속할 예측확률을 추정한다. 이러한 통합문화이용권 가맹점 데이터에 내재한 패턴을 토대로 하여 문화예술위원회가 사업자를 가맹점으로 지정하거나 기존 가맹점을 해지하는 의사결정을 설명하는 모형을 구성하였다. 기본 모형으로서 문화예술위원회가 사업자에 대한 의사결정을 ‘가맹점 승인/가맹점 해지’로 이진 분류하여 가맹점 승인 예측확률을 계산하는 모형을 구성하였다.

[표 2] 분석에 사용된 특성 및 범주

[Table 2] List of features and classification used in analysis

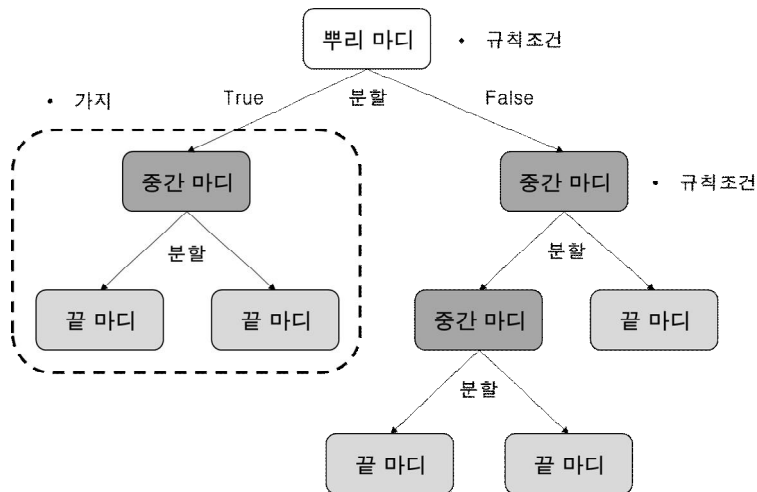
특성		범주
승인여부		승인, 해지
구분		[표 1]의 대분류
판매형태 구분		온라인, 오프라인, 온/오프라인
구분(중분류)		[표 1]의 중분류(장르)
구분(소분류)		[표 1]의 소분류
사유 및 기타	단말기 분리	영업장 내 비허용품목과 허용품목을 결제하는 단말기가 분리되어 있는 경우
	비허용품목	영업장 내 비허용품목이 있는 경우
	비허용비율	영업장 내 비허용품목 판매 비율이 10% 이하인 경우
	해지요청	문예위 담당자가 해지요청을 한 경우
	승인요청	문예위 담당자가 승인요청을 한 경우
	한시적	사업자를 축제 등 점포를 한시적으로 운영한 경우
	재등록중	다양한 사유에 따라 사업자를 가맹점으로 재등록중인 경우
	재등록예정	해당 사업자가 가맹점으로 재등록할 예정인 경우
	복수	복수 가맹점번호 사용 및 비인증 가맹점번호로 비공개 추가 등록
	폐업확인	해당 영업장 폐업
	협력발굴건	사업자를 협력사업으로 발굴한 경우

문화예술위원회의 의사결정을 설명하기 위한 변수는 「2020년 통합문화이용권 사업지침」의 내용과 관리데이터 상의 가맹점 정보를 참고하여 선정하였다. 문화체육관광부는 매년 통합문화이용권 사업지침을 작성하여 배포해오고 있는데, 여기에는 통합문화이용권 가맹점 가능 업종과 예외적

허용지침을 제시하고 있다. 이러한 정보를 데이터셋의 특성(feature)으로 변환 및 추가하여 특성 조합과 가맹점 승인 관련 의사결정 간의 관계 규칙을 규명한다. 모형에서 사용된 특성을 나열하자면 가맹점 분류정보, 2020년 결제 건수, 결제 승인금액 정보, 가맹점 관리데이터에 기록된 등록 사유 및 데이터베이스에 기록된 텍스트 데이터를 코딩한 정보로 요약할 수 있고, [표 2]에 정리하였다. 가맹점 관리데이터에 기록된 등록 사유 및 기타 정보는 텍스트 데이터를 내용 분석(content analysis)하여 이를 개방 코딩(open coding)형태로 범주화하였다. 즉, 텍스트 데이터 내에 정보가 유사하거나 의미상 관련된 사고, 사건, 물체, 작용/상호작용 등을 하위범주(subcategory)로 분류한 후 범주화하는 질적방법론(qualitative method)의 방법인 개방 코딩을 시행했다 [29]. 하위범주로 범주화하는 과정에서는 「2020년 통합문화이용권 가맹점 관리지침」에 기재된 예외 사항을 주로 참고하였고 그 외에도 가맹점 관측치에 공통으로 나타난 키워드를 선정하여 이를 특성으로 변환하였다.

분석을 위해서 분류 예측 기계학습에서 의사결정나무(Decision Tree) 모형을 사용하였다. 의사결정나무 모형은 데이터에 존재하는 패턴을 나무 형태로 도표화하여 분류 또는 수치를 예측하는 단일모형으로 정의된다 [30]. 의사결정나무는 계층적 구조(hierarchical structure)를 취하기 때문에 예측이 어떻게 이루어지는지를 비교적 명확히 이해할 수 있다는 장점이 있다. 즉, 모형이 예측하는 방식을 분석자가 직접 확인이 가능하다는 점에서 전형적인 화이트박스(white box) 모형 형태를 취하고 있으므로 시각화와 해석이 쉽다. 또한, 모형을 생성하기 위한 데이터 전처리 과정이 다른 모형보다 상대적으로 적어서 효율적이고, 모형 스스로 중요한 변수를 선택하는 특성 선택기능(feature selection)을 포함하고 있고 편의성을 지니고 있기도 하다. 구체적으로 의사결정나무는 최적의 분할 변수와 분할점을 선택하는 것을 목표로 하고, 비슷한 범주를 갖는 관측치끼리 나무의 마디를 형성하고 뿔어나가는 식으로 시각화할 수 있다. [그림 1]은 의사결정나무 모형의 구성요소를 그림으로 나타낸 것이다.

[그림 1]에 나타난 의사결정트리는 if-then-else 규칙으로 구성된 집합체라고 볼 수 있고, 의사결정나무 모형은 뿌리마디, 끝마디, 가지로 구성되어 있다. 여기서 뿌리마디(root node)는 분할이 시작되는 시점으로서 첫 번째 분할과 관련한 규칙을 명시하고 있다. 다음으로 중간마디(internal node)는 뿌리 마디로부터 최종 마디 전까지 분할된 부분을 표시하고 여기에서 하위 마디로 뿔어나가기 위한 분할 규칙을 명시하고 있다. 끝마디(terminal node)는 기계가 최종적으로 결정한 예측값을 기입하는데, 나무의 구조에 빗대면 잎사귀(leaf)에 해당된다. 끝으로, 가지(branch)란 하나의 마디부터 끝마디까지 연결된 마디의 집합을 의미한다. 본 연구의 분석에서는 목표변수의 데이터 유형이 ‘승인’과 ‘해지’ 범주로 구성되어 있으므로 분류나무(classification tree) 모형을 구성하여 끝마디에서 각 클래스의 비율을 비교하고, 이를 토대로 가장 높은 비율을 갖는 클래스를 결정하여 종속변수를 예측하는 방식을 채택한다.



[그림 1] 의사결정나무 모형의 구성요소

[Fig. 1] Components of Decision Tree Model

의사결정나무의 정확도 성능을 평가하기 위해 평균 정확도 감소량(Mean Decrease Accuracy, MDA)계산을 통해 규칙조건 내 사용한 속성을 모형에서 제거할 때 정확도가 얼마나 감소하는지를 계산하여 정확도 개선에서 가장 중요한 변수를 찾는다. 또한, 의사결정나무는 각 마디에서 불순도를 측정하여 불순도가 낮아지는 방향으로 마디를 형성해 나가게 된다. 이때 불순도(impurity)란 특정 마디에 얼마나 다양한 범주가 포함되어 있는지를 나타내는 지수를 의미한다. 불순도를 측정 및 판별하기 위해 분석에 지니지수(Gini Index)를 사용하게 되는데, 본 연구에서는 평균 지니 불순도 감소량(Mean Decrease Gini, MDG) 계산을 통해 속성을 모형에서 제거할 때 불순도를 얼마나 감소시키는지를 측정함으로써 변수 중요도(feature importance)를 측정하였다.

의사결정나무 모형의 최종 평가는 예측 의사결정과 실제 의사결정에 관한 오차 행렬(confusion matrix)을 구성하고, 이를 토대로 정확도, 정밀도, 재현율, F1-Score를 계산함으로써 진행한다 [31]. 이러한 성능평가지수를 계산하여 기계에 훈련데이터를 학습시켜서 구성한 모형의 예측값과 실제 값을 비교할 수 있다. 우선, 정확도(accuracy)는 모형이 얼마나 바르게 분류하는지를 측정하는 비율로 정의되고, 정밀도(precision)는 모형이 True(승인)으로 분류한 것 중 실제 값이 승인인 비율로 정의된다. 재현도(recall)는 실제값이 True(승인)인 것 중 모형이 승인으로 분류한 비율로 정의된다. F1-Score는 정밀도와 재현도의 조화평균(harmonic mean)을 의미하고 데이터가 특정 분류 클래스로 치우쳐져 있는 불균형을 이룰 때 정확도 대신 쓰인다.

의사결정 분석에서 모형의 가지 수는 분류가 유의미한 수준까지만 형성되도록 가지치기(pruning)를 수행하였다. 즉, 뿌리마디가 (승인, 승인) 또는 (해지, 해지) 형태로 나타나는 가지는 분석의 관심사가 아니기 때문에 이를 불필요한 의사결정 포인트로 간주하여 가지를 제거한다.

## 2.2 가맹점 분류 분석 결과

### 2.2.1 기술통계 분석

가맹점 승인 및 해지 결정 분포를 살펴본 결과 승인 비율이 60.8%(25,924개)를 차지하고 해지 비율이 37.9%(16,153개)를 차지하는 것으로 나타났다. 승인 비율이 해지 비율보다 높은 경향을 보이나 약 6:4 정도의 비율이므로 이진 분류 모형 상 불균형으로 분류될 수준은 아님을 알 수 있다. 대부분의 가맹점은 오프라인 형태로 분포되어 있고, 온라인 형태의 가맹점은 3.6%(1,538개)를 차지하고 있다. 대분류 기준에 따라 살펴보면 문화(47.4%), 관광(37.3%), 체육(15.3%) 순으로 가맹점 비중이 나타난다. 중분류 기준에 따라 살펴보면 숙박(19.3%), 미술(16.5%), 도서(14.4%)로 분류되는 가맹점이 전체 가맹점의 약 절반을 차지하고 있다. 소분류 기준으로는 숙박(19.3%), 도서(14.4%), 사진관(11.9%), 체육시설(8.9%)로 분류되는 가맹점이 전체 가맹점의 절반 이상을 차지한다. 가맹점 결제 데이터에 있는 가맹점의 장르 분포를 보면 도서(22.6%), 숙박(17.2%), 사진관(12.9%)이 전체 가맹점의 약 절반을 차지하고 있다.

[표 3]은 가맹점 결제기록에 관한 기술통계량을 요약하고 있다. 가맹점의 결제승인금액은 평균값이 중간값보다 큰 양의 비대칭 분포(positively skewed distribution)를 보인다. 통합문화이용권 가맹점은 평균 8,658,837원의 결제승인금액을 기록하였다. 중간값은 585,520원으로서 평균값과 다소 차이를 보인다. 가장 많은 결제승인금액을 기록한 가맹점은 교보문고(도서)이고, 한국철도공사(철도), 영풍문고(도서), 현장발권-S(시외/고속버스), 전국버스운송사업조합연합회(시외/고속버스) 순으로 나타났다.

[표 3] 가맹점 결제기록 기술통계량

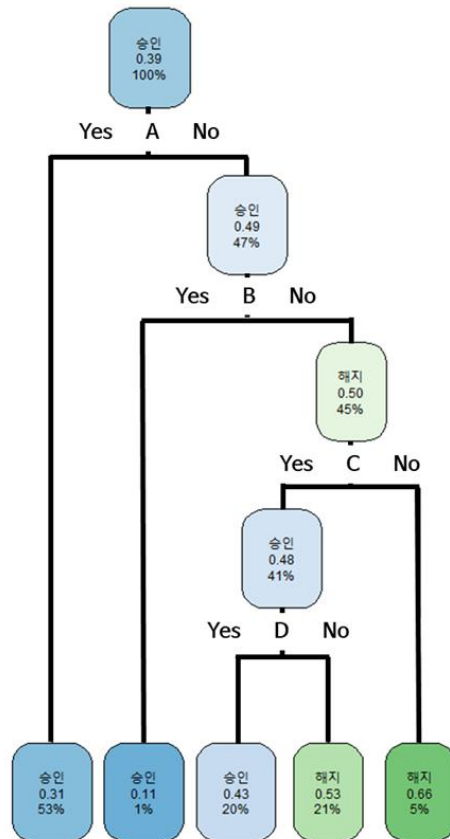
[Table 3] Descriptive statistics of affiliate payment

	결제승인금액(단위: 원)	결제승인건수(단위: 건)
최솟값	10	1
최댓값	10,181,071,180	473,988
중간값	585,520	15
표준편차 (분산)	109,427,060.3 (11,974,281,520,000,000)	6186.35 (38,270,937.5)
변동계수	12.64	15.81
왜도	5188.78	3662.3
첨도	63.68	57
평균	8,658,836.6	391.26
중앙값 절대편차 (Median Absolute Deviation)	505,520	12,659



2020년 가맹점의 결제승인건수는 평균값이 중간값보다 큰 양의 비대칭(positive skewness) 분포를 보인다. 2020년 가맹점은 평균적으로 391건의 결제승인건수를 기록하였고, 중간값은 15건으로 평균 결제승인 건수와 다소 차이를 보인다. 가장 많은 결제 건수를 기록한 가맹점은 한국철도공사(철도)이고, 교보문고(도서), 현장발권-S(시외/고속버스), 전국버스운송사업조합연합회(시외/고속버스), 영풍문고(도서) 순으로 나타났다.

### 2.2.2 의사결정나무 모형 분석 결과

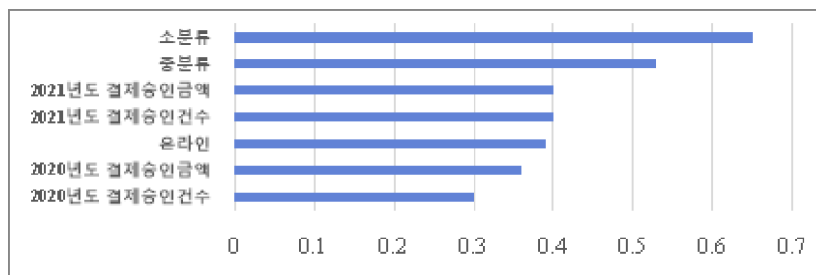


[그림 2] 통합문화이용권 가맹점 승인/해지 의사결정나무 모형

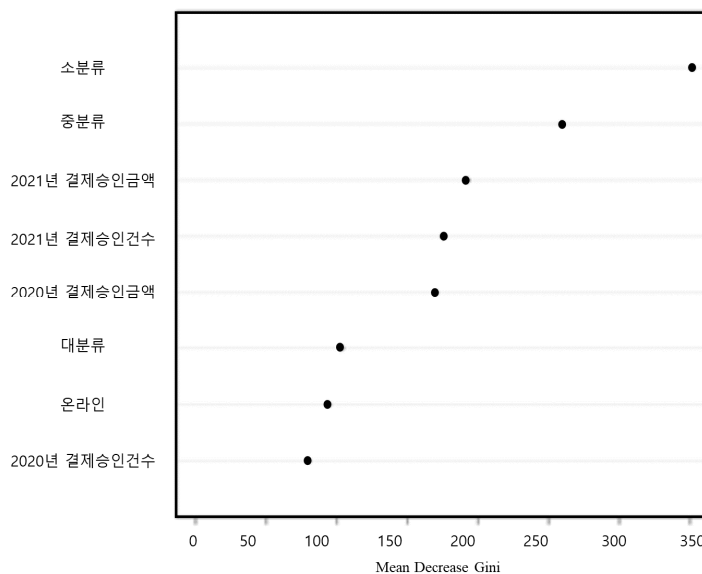
[Fig. 2] Decision Tree Model of Affiliate

가맹점 결제 이력 데이터를 의사결정나무로 분석한 결과는 시각화할 수 있다. [그림 2] 끝 마디를 살펴보면 전체 100% 데이터 중 승인과 해지의 비율은 7.4:2.6의 비율로 나타난다. 첫 번째 분기 조건에서 A조건(소분류 상으로 TV, 공연, 공예, 관광명소, 국내항공, 도서, 동·식물원, 문화체험, 스포츠 관람, 시외/고속버스, 여객선, 영화, 온천, 음악, 직업체험, 철도, 체육시설, 체육용품, 테마파크)에 해당하는지 검사하여 True면 승인(왼쪽 노드)으로 분류하는데, 이는 전체 데이터의 53%를 차지

한다. 두 번째 분기조건에서는 B조건(온라인 판매자)에 해당하는지를 검사하여 True면 승인(왼쪽 노드)으로 분류하고 False면 해지(오른쪽 노드)로 분류하는데, 이는 각각 전체 데이터의 2%와 45%를 차지한다. 세 번째 분기조건에서는 C조건(2020년도 결제승인금액이 49,000원 이상)에 해당하는지를 검사하여 True면 승인(왼쪽 노드)으로 분류하고 False면 해지(오른쪽 노드)로 분류하는데, 이는 각각 전체 데이터의 41%와 5%를 차지한다. 끝으로, 네 번째 분기조건에서는 D조건(소분류 상으로 렌터카, 사진관, 여행사, 전시, 휴양림 · 캠핑장)에 해당하는지를 검사하여 True면 승인(왼쪽 노드)으로 분류하고 False면 해지(오른쪽 노드)로 분류하는데, 이는 각각 전체 데이터의 20%와 21%를 차지한다.



[그림 3] 의사결정나무 평균 정확도 감소량  
[Fig. 3] Mean Decrease Average of decision tree model



[그림 4] 의사결정나무 평균 지니 감소도  
[Fig. 4] Mean Decrease Gini of decision tree model

[그림 2]의 의사결정나무 모형에서 사용된 변수들의 중요도 분석결과를 [그림 3]의 MDA와 [그림 4]의 MDG를 통해 시각화할 수 있다. MDA 계산 결과 평균 정확도 감소량 계산 결과 소분류 기준(숙박, 체육시설, 영화), 2021년도 결제 승인금액, 중분류 기준(여행사), 2021년도 결제승인건수 순으로 가장 중요한 속성으로 나타났다. MDG 계산 결과 소분류, 중분류, 2021년도 결제승인금액, 2021년도 결제승인 건수, 2020년도 결제승인금액, 대분류, 온/오프라인, 2020년도 결제승인 건수 순으로 가장 중요한 속성이 나타났다.

의사결정나무 모형 결과 얻은 성능 평가 지수는 [표 4]와 같이 정확도 0.64, 정밀도 0.66, 재현도 0.85, F1-Score 0.74로 나타났다. 즉, 의사결정나무의 예측은 정답과 64% 정도 정확한 것으로 나타났고, 예측한 것 중에 정답의 비율이 66%였고, True나 False로 찾아야 할 의사결정을 실제로 찾은 비율은 85%에 해당하며, 정밀도와 재현율의 평균은 74% 수준으로 해석할 수 있다.

[표 4] 의사결정나무 모형 평가지표

[Table 4] Evaluation metric of decision tree model

	측정결과
정확도(Accuracy)	0.64
정밀도(Precision)	0.66
재현도(Recall)	0.85
F1-Score	0.74

### 3. 분석결과

가맹점 결제 데이터를 분석한 결과 가맹점 승인/해지 의사결정에 영향을 미치는 주요 변수를 발견하였다. 첫째, 대분류나 중분류 기준보다 소분류 기준이 상대적으로 의사결정에 더 많은 영향을 미치는 것으로 확인되었다. 분류 기계학습으로 구성된 의사결정모형은 가맹점 관리지침에서 제시하는 방향(소분류 기준을 신설하여 기준을 세분화)을 학습한 것으로 해석할 수 있다. 둘째, 결제 승인 건수와 승인 금액이 의사결정에 영향을 미치는 것으로 나타났다. 물론, 빅데이터 분석이 인과관계가 아닌 상관관계를 설명하므로 결제 빈도와 결제 규모가 작을수록 해지 확률이 증가하는지 또는 해지된 가맹점일수록 결제 빈도와 결제 규모가 적은지를 명확히 설명하기는 어려운 측면이 있다. 셋째, 온라인 가맹점에 관한 의사결정은 오프라인 가맹점보다 상대적으로 단순한 패턴을 보인다. 이는 온라인 판매자의 판매 품목 정보가 오프라인 가맹점의 품목 정보보다 비교적 투명하게 공개되어 있다는 특성을 반영한 것으로 판단된다. 넷째, 관광(대분류)-숙박(중분류)-숙박(소분류)으로 분류되는 가맹점에 대한 의사결정이 가장 복잡한 것으로 나타났다. 즉, 소분류 상으로 교통수단, 미술, 스포츠 관람, 렌터카, 사진관, 여행사, 전시, 휴양림/캠핑장으로 분류되는 가맹점에 대한 의사결정 또한 다양한 변수가 영향을 미치는 것으로 추측되고, 단말기 분리 여부, 재등록 과정 중

여부, 복수 가맹점 여부, 문화예술위원회 담당자의 승인요청 사유에 따라 승인과 해지 의사결정에 영향을 미치는 것으로 보인다.

#### 4. 결론

본 연구는 통합문화이용권에서 파생된 2020년, 2021년 가맹점 관리 빅데이터를 기반으로 이용자의 문화누리카드 사용패턴을 탐색하기 위해 의사나무결정 분석을 수행하였다. 이러한 사용패턴 분석을 통해 이용자의 욕구와 트렌드를 반영할 수 있는 실질적인 통합문화이용권 관리 방안을 모색하려 하였다.

분석을 토대로 다음과 같은 가맹점 관리 관련 시사점을 제시할 수 있다. 첫째, 현행 관리체계에서는 관광 장르의 가맹점 관리비용이 문화와 체육 장르의 관리비용보다 더 많이 소요되는 것으로 추정된다. 그러나 이러한 추정결과에 대한 명확한 결론 도출을 위해서는 관광장르 가맹점에 대한 관리 규정(예: 복합사업자 세부사항 등)이나 통합문화이용권 운영과정에 대한 심층분석을 통해 가맹점 관리상의 변동요인(승인/해지 요인)을 규명하여 관리적 효율성을 개선하는 것이 필요할 것이다. 관광 장르 이외에도 문화와 체육 장르 중 해지빈도가 높은 업종(예: 문화-미술-사진관 장르 등)에 대해서 관리상의 변동요인을 규명하여 관리상의 효율성을 도모하는 것이 필요하다. 즉, 가맹점 관리 규정을 더욱 엄격하게 하여 부실 가맹점을 과감히 정리하고 신규 가맹점을 까다롭게 관리하거나, 규정을 더욱 완화하여(허용 업종 및 품목 확대 등) 가맹점 이용 편의성을 확대하거나, 문제가 되는 업종 자체를 통합문화이용권에서 제외하는 대안 등을 고려해 볼 수 있다.

둘째, 가맹점 관리체계가 세분화될수록 가맹점 관리를 위한 의사결정 기준이 명확해져서 의사결정 비용(승인/해지 여부 결정)이 감소하는 경향이 있다. 따라서, 현행 “대분류-중분류-소분류” 체제에서 각 분류 단계를 좀 더 섬세하게 세분화하거나 “대분류-중분류-소분류-세분류” 등으로 하위 분류를 신설하는 방안을 고려하는 것이 바람직해 보인다. 가맹점 데이터 입력에 있어 자주 등장하는 관리상의 사유는 자연어 형태가 아닌 카테고리 방식으로 분류하는 것이 바람직하다. 예를 들어, 분석에서 가공·활용된 방식으로 승인/해지 사유를 저장하여 데이터 관리상의 효율성과 분석상의 체계성을 향상시킬 수 있을 것이다.

#### References

- [1] H. S. Yang, D. W. Lim, H. G. Z, “Personal Information Overload and User Resistance in the Big Data Age”, *Bibliographic Info (Intell Inform Syst)*, vol. 19, no. 5, March 2013, pp. 125-139.
- [2] S. Sagioglu, D. Sinanc, “Big data: A review”, 2013 International Conference on Collaboration Technologies and Systems (CTS), May 20-24, 2013, San Diego, CA, USA, pp. 42-47, doi: 10.1109/CTS.2013.6567202.

- [3] J. S. Kim, "Big data analysis Technologies and practical examples", Review of Korea Contents Association, vol. 12, no. 1, March 2014. pp. 14-20. doi: 10.5626/KTCP.2020.26.8.349.
- [4] K. R. Jung, "Utilizing Big Data In Arts & Culture Welfare Services", The Korean Journal of Arts Studies, vol. 25, September 2019, pp. 5-28. doi: 10.20976/kjas.2019..25.001.
- [5] S. M. Chang, Y. H. Lee, "A Study of Consumer Perception on Opera Performances Using Big Data", The Journal of Culture Contents, vol. 23, December 2021, pp. 87-115. doi: 10.34227/tjocc.2021..23.87.
- [6] Y. G. Lee, H. C. Joo, "A Study on the Recognition of Dance Performance Using Big Data Analysis", The Korean Journal of Sport, vol. 18, no. 2, June 2020. pp. 365-374. doi: 10.46669/kss.2020.18.2.034.
- [7] J. S. Kang, B. A. Rhee, "A Visitor Study of The Exhibition of <David Hockney> Using Big Data Analysis which reflects viewing experiences", Journal of The Korea Society of Computer and Information, vol. 27, no. 2, February 2022, pp.81-89. doi: 10.9708/jksoci.2022.27.02.081.
- [8] Y. J. Lee, J. J. Koo, "Visualizing Public Perception of Art Rental by Big Data Analysis", Bulletin of Korean Society of Basic Design & Art, vol. 22, no. 3, June 2021, pp. 373-384. doi: 10.47294/KSBDA.22.3.28.
- [9] D. Y. Jang, J. Y. Kim, "A study on the Responses Analysis of Viewers of Superhero Films Based on Their Online Reviews from the Big Data", Design convergence study, vol. 20. no. 2, April, 2021, pp. 39-57. doi: 10.31678/SDC87.3.
- [10] S. Y. Lee, H. J. Yoon. "A Study on Big Data Information System based on Artificial Intelligence -Filmmaker and Focusing on Movie case analysis of 10 million Viewers", The Korea Institute of Electronic Communication Sciences, vol. 14. no. 2, April 2019, pp. 377-388. doi: 10.13067/JKIECS.2019.14.2.377
- [11] J. Y. Kwon, Y. S. Oh, K. K. Choi, "Big Data Analysis on Out-of-School Youth Arts and Culture Education", The Korean Society of Sports Science, vol. 29, no. 5, October 2020, pp. 857-870. doi: 10.35159/kjss.2020.10.29.5.857.
- [12] J. S. Lee, J. Y. Park, D. G. Kim, "Exploring the Last of University Students' Cultural and Artistic Enjoyment by Analyzing Text Mining", The Journal of Humanities and Social science. vol. 11, no. 6, December 2020, pp. 1821-1834. doi: 10.22143/HSS21.11.6.129.
- [13] D. W. Park, I. S. Koh, N. S. Lee, K. S. Han, "A Study on Architecture for BigData-based Book Curation System", Journal of the Korea Society of Information Technology Policy & Management (ITPM), vol. 12, no. 1, December 2020, pp. 1559-1565.
- [14] D. H. Kim, S. K. Lee, "A Study on the Strategies Used from 4th Industrial Core Technology for Activating Culture in the Islands Area", The Journal of Korean Island, vol. 30, no. 4, December 2018, pp. 71-96. doi: 10.26840/JKI.30.4.71.
- [15] M. H. Lee, I. W. Jeon, C. M. Jun, "Demand Estimation of Public Transport using Smart Card Big Data", Korea Spatial Information Society (KSIS), vol. 28, no. 3, September 2020, pp. 3-10. doi: 10.7319/kogsis.2020.28.3.003.
- [16] Y. C. Kim, C. S. Kim, S. B. Kim, "An Algorithm for Extracting Tourists' O-D Patterns Using Encrypted Smart Card Data of Public Transportation: Application to Tourist City, Jeju", Journal of KISS(Computing Practices), vol. 26, no. 8, August 2020, pp. 349-361, doi: 10.5626/KTCP.2020.26.8.349.

- [17] T. C. Jin, E. H. Jeoung, “Big Data Analysis of Social Media on Gangwon-do Tourism”, *Journal of Korea Institute of Information, Electronics, and Communication Technology*, vol. 14, no. 3, June 2021, pp. 193-200. doi:10.17661/JKIIECT.2021.14.3.193.
- [18] Y. I. Yoon, J. J. Choi, “Analyzing the Sustainable Factors of Attractiveness at Tourism Destination of Film Locations Through Social Big Data”, *Journal of Tourism Sciences*, vol. 42, no. 6, June 2018, pp. 63-79. doi: 10.17086/JTS.2018.42.6.63.79.
- [19] Y. I. Yoon, D. H. Ha, “How Business Hotel is Perceived with Big Data Analytics of Social Media”, *Management Education Review*, vol. 32, no. 3, June 2017, pp. 379-398. doi: 10.23839/kabe.2017.32.3.379.
- [20] J. H. Han, E. J. Noh, “The effect of price on customer review ratings between budget hotel and guest house: Online review big data analysis”, *Journal of Tourism Management Research*, vol. 88, March 2019, pp. 331-352. doi: 10.18604/tmro.2019.23.2.16.
- [21] G. T. Kim, “Professional Baseball Viewing Culture Survey According to Corona 19 using Social Network Big Data”, *Journal of Korea Entertainment Industry Association*, vol. 14, no. 6, August 2020, pp. 139-150. doi: 10.21184/jkeia.2020.8.14.6.139.
- [22] C. S. Lee, D. H. Kim, S. H. Hwang, “Consumer Sentiment of Korean Professional Baseball Spectators in Terms of Big Data”, *The Korean Journal of Sport*, vol. 17, no. 2, June 2019, pp. 881-889.
- [23] S. Y. Shin, J. H. Yang, “A Study on COVID-19 and Consumption Expenditure for Sports Services and Products: Trend Analysis by Business Category based on Credit Card Big-Data”, *Journal of Korea Service Management Society*, vol. 22, no. 4, November 2021, pp. 190-222. doi: 10.15706/jksms.2021.22.4.0.
- [24] M. S. Lee, K. H. Chung, J. H. Park, “Big Data Analysis of Global Sportswear Brands”, *Journal of the Korean Society of Costume*, vol. 71, no. 5, October 2021, pp. 96-116. doi: 10.7233/jksc.2021.71.5.096.
- [25] S. W. Oh, J. W. Han, M. S. Kim, “A Social Big Data Analysis on Perception about Water Leisure Sport Tourism”, *Korean Society For Sport Management*, vol. 24, no. 4, October 2019, pp. 83-95. doi: 10.31308/KSSM.24.4.6.
- [26] J. H. Han, “Analysis of Leisure Sports Consumers’ Perception of Public Golf Courses and Exploration of Future Facilities: Focused on Big Data”, *Journal of Tourism and Leisure Research*, vol. 33, no. 7, July 2021, pp. 395-410. doi: 10.31336/JTLR.2021.7.33.7.395.
- [27] Arts Council Korea, “Integrated Culture Voucher: Person in charge work manual”, Arts Council Korea, Naju, Republic of Korea, January 2022. [Online]. Available: <https://www.mnuri.kr/notice/data/detail.do>.
- [28] H. S. Kim, *Business Machine Learning in Python*, Predix, 2020.
- [29] K. Charmaz, *Constructing grounded theory: a practical guide through qualitative analysis (Second Edition)*, SAGE Publications LTD, 2014.
- [30] A. Géron, *Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow: Concepts, Tools, and Techniques to Build Intelligent Systems (Second Edition)*, O’Reilly Media, 2019.
- [31] H. S. Kim, *Business Machine Learning in Python*, Predix, 2020.