

머신러닝을 활용한 기상조건에 따른 공공도서관 도서대출 수요분석

오민기¹, 김건욱^{2*}, 신세영³, 이진명⁴, 장원준⁵

¹경북대학교 문헌정보학과 학부과정, ²대구디지털산업진흥원 빅데이터활용센터 센터장,
³경북대학교 컴퓨터학과 학부과정, ⁴경북대학교 물리학과 학부과정, ⁵대구디지털산업진흥원 전임연구원

Analysis of public library book loan demand according to weather conditions using machine learning

Min-Ki Oh¹, Keun-Wook Kim^{2*}, Se-Young Shin³, Jin-Myeong Lee⁴, Won-Jun Jang⁵

¹Undergraduate, Dept. of Library and Information Science, Kyungpook National University

²Director, Big Data Center, Daegu Digital Industry Promotion Agency

³Undergraduate, Dept. of Computer Science, Kyungpook National University

⁴Undergraduate, Dept. of Physics, Kyungpook National University

⁵Associate Research Engineer, Big Data Center, Daegu Digital Industry Promotion Agency

요약 국내 공공도서관은 1, 2차 도서관 발전 종합계획을 토대로 양적 성장을 이루었으나, 질적으로는 다소 부족한 점이 있어 이를 개선하기 위한 다양한 연구가 수행되었다. 대다수 선행연구에서는 사회·경제적 요인과 통계분석에 한정되어 수행된 한계점이 있다. 이에 본 연구에서는 시공간적 개념을 적용하여 강우와 폭염으로 인한 공공도서관 대출 수요 감소를 정량적으로 산출하고, 기상 변화로 도서 대출 수요 감소가 높은 지역과 그렇지 않은 지역을 군집화하여 공공도서관 내·외부 요인들과 결합한 후 기상변화에 따른 공공도서관 대출 수요 변화를 분석하였다. 분석 결과 공공도서관별 기상으로 인한 감소 차이가 존재하였으며, 공공도서관의 특성과 공간적 위치에 따라 일부 다르게 나타났다. 또한, 기온이 35°C 이상인 폭염일 경우 도서 대출 수요 감소 폭이 많이 증가하였으며, 랜덤포레스트 모형으로 분석한 결과 유의미한 요인이 도출되었다. 내적 요인으로는 좌석 수, 장서 수, 면적이 도출되었으며, 외적 요인으로는 공공도서관 접근 경사로, 카페, 독서실, 10대 유동인구, 30/40대 여성 유동인구가 중요한 변수로 분석되었다. 이러한 분석 결과는 특정 시즌 기상을 고려한 공공도서관 이용 활성화 정책 수립에 이바지할 것으로 판단되며, 연구의 한계점도 제시하였다.

주제어 : 공공도서관, 기상, 요인분석, 랜덤포레스트, 머신러닝

Abstract Although domestic public libraries achieved quantitative growth based on the 1st and 2nd comprehensive library development plans, there were some qualitative shortcomings, and various studies have been conducted to improve them. Most of the preceding studies have limitations in that they are limited to social and economic factors and statistical analysis. Therefore, in this study, by applying the spatiotemporal concept to quantitatively calculate the decrease in public library loan demand due to rainfall and heatwave, by clustering areas with high demand for book loan due to weather changes and areas where it is not, factors inside and outside public libraries and After the combination, changes in public library loan demand according to weather changes were analyzed. As a result of the analysis, there was a difference in the decrease due to the weather for each public library, and it was found that there were some differences depending on the characteristics and spatial location of the public library. Also, when the temperature was over 35°C, the decrease in book loan demand increased significantly. As internal factors, the number of seats, the number of books, and area were derived. As external factors, the public library access ramp, cafe, reading room, floating population in their teens, and floating population of women in their 30s/40s were analyzed as important variables. The results of this analysis are judged to contribute to the establishment of policies to promote the use of public libraries in consideration of the weather in a specific season, and also suggested limitations of the study.

Key Words : Public library, Weather, Factor Analysis, Random Forest, Machine Learning

*Corresponding Author : Keun-Wook Kim(aut7767@dip.or.kr)

Received December 28, 2021

Revised February 4, 2022

Accepted March 20, 2022

Published March 28, 2022

1. 서론

공공도서관은 도서관 자료와 서비스를 지역주민들에게 제공하여 지적 요구 충족, 평생학습 등에 이바지하는 시설을 의미하며, 사회환경 변화에 따라 복합문화공간, 지역 커뮤니티, 정보격차 해소의 역할을 수행하고 있으며, 그 중요성은 지속해서 증가하고 있다.

국내 공공도서관은 국민의 지적 수요에 발맞춰 지식 정보 강국으로 발전하기 위해 1, 2차 도서관 발전 종합계획을 토대로 공공도서관의 지속적인 확충과 성장을 추진 과제로 성장하여 2018년 기준 공공도서관 1,096관으로 양적 성장을 달성하였고, 거주민들의 공공도서관 접근성은 향상되었다[1, 2].

공공도서관의 양적 성장에도 불구하고 미 대출 도서의 증가, 방문자 수의 감소, 도서관 이용 형태의 변화 등 질적 개선에 있어 일부 한계가 존재하며, 이로 인해 공공도서관 이용 활성화를 위한 다양한 연구가 진행되었다[3].

대다수 선행연구에서는 공공도서관 내적 요인인 장서 수, 프로그램 수, 규모 등의 변수와 외적 요인인 접근성, 인구 특성, 경제적 여건 등의 변수를 활용하여 공공도서관 이용 활성화 정책 수립에 이바지하였으나, 대다수 통계 분석과 사회·경제학 변수에 한정된 분석이 주를 이루는 한계점을 가지고 있다[4-7].

한편 최근 지구 온난화로 기상 변화가 급격히 이루어져, 기상과 공공시설물 이용과 관련된 연구가 활성화되고 있으나, 기상과 공공도서관과의 관계를 고려한 연구는 부재하며, 공공에서는 대다수 교통 분야에 국한되어 연구가 수행됐다[8-12].

이러한 배경 하에 본 연구는 대구시 공공도서관 대출 이력 자료와 기상청 중간기상관측(ASOS) 데이터를 활용하여, 기상이 공공도서관 이용에 미치는 영향 정도를 파악하고, 머신러닝 기반의 변수 중요도를 도출하는 것을 목적으로 한다. 본 연구의 분석 결과는 기상과 공공도서관과의 관계 연구에 있어 기초자료로 활용될 것으로 판단되며, 특정 시점 기상을 고려한 공공도서관 이용 활성화 정책 수립에 이바지할 것으로 기대한다.

2. 선행연구 고찰

공공도서관 관련 선행연구들은 도서 대출 수요와 이용자의 행태와 관련된 연구들이 주를 이루고 있으며, 유의미한 영향을 미치는 요인으로는 외적 요인과 내적 요인

으로 구분할 수 있다[4-7, 13-15].

공공도서관 외부에서 발생하는 외적 요인으로는 지리적, 인구·사회학적, 교통 및 환경 요인들로 구성되며 [4-6], 내적 요인으로는 도서관 규모, 자료, 이용자 등이 포함된다[7].

박진규, 김인(2016)은 공공도서관의 입지 형태에 따른 거리 중심의 지리적 접근성을 바탕으로 도서관 이용자의 이용 패턴과의 관계를 정량적으로 분석하였고, 공공도서관과 거주지, 버스 정류장까지의 거리가 주된 요인으로 분석되었다[4].

전계형, 권성영(2018)은 패널 고정효과 모형으로 지역의 인구 및 경제적 상황의 변화가 도서관 이용에 미치는 영향을 확인하였고, 분석 결과 경제적 환경 변화가 도서관 대출자 수에 주요한 요인으로 나타났다[5].

박형준, 전선민, 정주철(2018)은 순서형 로짓모형을 활용하여 공공도서관의 이용 만족도를 대중교통 시설과의 접근성, 인근 공원과의 연계 여부 등 도시 계획적 측면에서 분석을 시도하였고, 분석 결과 인근 공원과의 연계 여부가 도서관 이용 만족도에 유의미한 영향 요인으로 나타났다[6].

이경진(2020)은 로그 중 회귀모형으로 도서관 방문자 수와 자료실 이용자 수에 영향을 주는 요인을 분석하였고, 도서관의 소장자료 규모, 시설 규모, 도서관 프로그램 및 교육과 운영, 이용자 활동이 유의미한 변수로 나타났다[7].

이성신(2011)은 공공도서관 자료 선정 및 서비스 개발에 지리정보시스템을 활용하여 상업적 관점에서 접근하였고, 분석 결과를 도서 이용 활성화를 위한 마케팅에 활용하는 방안을 제시하였다[13].

김태영, 백지연, 오효정(2018)은 빅데이터 로그를 활용하여 도서관 이용자 및 대출 현황을 분석하고, FGI(Focus Group Interview)와 외부 데이터와 연계 분석하여 도서관 이용자 및 대출 정보 분석 결과에 대해 요인분석을 하였으며, 그 결과 도서관 이용 연령별, 거주지별, 월별 연관성을 발견하였다[14].

이상의 선행연구들은 Table 1과 같으며, 대다수 선행연구에서 공공도서관 내·외적인 변수들을 기반으로 이용 활성화를 위한 연구가 수행된 장점이 있으나, 기상과 관련된 연구는 부재한 상황이다. 이에 본 연구에서는 기상과 교통과의 관계를 연구한 Chung[9]의 분석 방법론을 참고하여 기상과 공공도서관 도서 대출 수요와의 관계를 분석하고자 한다.

이러한 분석 결과는 강우와 폭염으로 인해 도서 대출

수요 감소가 상대적으로 높은 지역과 낮은 지역을 분류하고, 내·외부적 다양한 요인들을 결합하여 영향을 미치는 요인을 도출하고자 한다. 이러한 분석 결과는 공공도서관 이용 활성화 정책 수립과 공공도서관 인건 인프라 정비 등의 기초자료로 활용될 것으로 판단된다.

Table 1. Prior Research on Factors Affecting Public Library Use

Author (Year)	Purpose of Study	Method	Variables
J. K. Park (2016)	An Analysis of Geographical Accessibility to Public Libraries in Metropolitan Area and Related Implications	Distance-Oriented Analysis	Accessibility(Bus Stop, Distance to Library and Residence)
G. H. Jeon (2018)	A Study on the Influence of Economic Factors on Library Use	Panel Fixed-Effect Model	Economic Circumstance
H. J. Park (2018)	The impact of public facilities on the perception of parents visiting public libraries	Ordinal Logistic Regression	Accessibility(Public Transportation) Linkage (Nearby Park)
K. J. Lee (2020)	The Influence Factors on the Numbers of Visitors and Reference Room Users of Public Libraries	Log-Log Multiple regression Model	Library Characteristics(Scale of Facility and Collection, Program, Education, Operation)
S. S. Lee (2011)	Applications of Geographic Information Systems in Public Library Marketing	Literature Review Method	-
T. Y. Kim (2018)	An Analysis of Library User and Circulation Status based on Bigdata Logs	Log Analysis, Interview	-

3. 공공도서관 대출 수요 및 기상 분석

3.1 데이터의 수집

본 연구는 기상 조건에 따른 공공도서관 도서 대출 수요 감소를 분석하기 위해 공간적으로는 대구시 공공도서관 36개관을 대상으로 하였으며, 시간적으로는 2018년 1월부터 2020년 6월까지 약 2년 6개월간의 공공도서관 대출 이력 데이터를 대구 빅데이터활용센터에서 연구 목적으로 자료를 수집하였다.

수집된 데이터는 도서 데이터와 기상 데이터로 도서 데이터는 비식별 처리되어, 대출 건별 이용자 아이디, 도서 아이디, 대출일, 반납일, 관리 도서관 등을 포함하고 있으며, 기상 데이터는 기상청 종관기상관측 데이터를

수집하여 지면온도, 5cm 지중온도, 10cm 지중온도, 기온, 강수량, 풍속, 적설 등의 정보를 포함하고 있다.

그 외에도 기상 변화에 따른 공공도서관 대출 수요 감소에 영향을 미치는 요인을 분석하기 위해 선행연구에서 유의미하다고 판단된 내·외부 요인들을 공공데이터와 민간데이터로부터 수집하여 최종적으로 분석에 활용된 데이터 목록은 Table 2와 같다.

Table 2. Components of Data

Public Library	ASOS
User Key, Book Key, Loan Date, Return Date, Manage Library and etc.	Rainfall, Humidity, Ground Temperature, Temperature, Wind Speed and etc.
Public	Private
Commercial District Information, National Library Standard Data, Slope, Bus Stop, Subway, School	Flow Population (2018.01-2019.12), Card Sales (2018.01-2018.12)

3.2 데이터 전처리

수집된 데이터에 존재하는 이상치(Outlier)를 제거하기 위해 별도의 전처리를 수행하였으며, 세부적인 내용은 다음과 같다.

첫째, 도서관 자료실 운영시간 외의 대출 기록을 제거하였으며, 대구시 공공도서관 자료실 운영시간은 도서관 별로 다르나 공통적으로 평일은 9시부터 18시, 주말은 9시부터 17까지 운영하며, 이 외의 시간은 무인도서 대출 반납기 또는 야간 대출 서비스로 이용할 수 있기에 해당 데이터는 제거하였다.

둘째, 도서관 이용 규정을 벗어난 데이터는 이상치로 판단하여 제거하였으며, 한 예로 도서관에서 하루 최대 대출할 수 있는 권수는 10권이며, 노약자나 장애인의 경우 최대 15권을 대여할 수 있어, 이를 초과하는 데이터에 대해서는 이상치로 판단하여 제거하였다.

셋째, 2020년 2월부터 유행한 감염병인 코로나19로 인해 공공도서관 이용객들의 감소가 높았으며, 이로 인한 영향력을 최소화하고자 해당 기간 이후의 데이터를 제거하여, 시간적 범위는 2018년 1월부터 2020년 1월까지로 변경하여 전처리한 결과 도서관 24개관, 대출 이력 6,047,723건이 최종 분석 대상에 포함되었다.

마지막으로 폭염으로 인한 도서 대출 수요 변화를 파악하기 위해서는 폭염과 체감온도의 파생변수 생성이 필요하며, 기상청의 생활 기상지수 자료에서 제공하는 '세분화된 체감온도 설명자료'를 활용하여 체감온도¹⁾를 계

산하였으며, 체감온도가 33°C 이상이면 폭염으로 정의하여 이를 준용하였다. 세부적인 체감온도 계산식은 다음과 같으며, 습구온도와 상대습도, 건구 온도가 계산식에 활용된다.

$$Tw = Ta \arctan(0.151977 (RH + 8.313659)^{\frac{1}{2}}) + \arctan(Ta + RH) - \arctan(RH - 1.676331) + 0.00391838 (RH^{\frac{3}{2}}) \arctan(0.023101 RH) - 4.686035$$

$$\text{체감온도} = -0.24418 + 0.553991 Tw + 0.455346 Ta - 0.00217 (Tw^2) + 0.002782 Tw Ta + 3.5$$

(Ta : 건구온도, Tw : 습구온도, RH : 상대습도(%))

3.3 도서관 대출 수요 및 기상 특성 분석

3.3.1 월별, 요일별, 시간대별 도서 대출 수요 특성

공공도서관 도서 대출 수요 현황을 월별로 살펴보면 Fig. 1과 같이 나타나며, 월평균 31,317명의 이용자가 도서관을 이용했으며, 통상적으로 1월, 7월, 8월에 도서관 이용자가 높은 것으로 나타났는데, 이는 선행연구[14, 15]의 분석 결과와 동일하게 방학 기간 도서관 이용자 수가 크게 증가함을 알 수 있다. 또한, 3월에는 새 학기 시작에 맞추어 도서관 이용자 수가 다소 증가한 것으로 보인다.

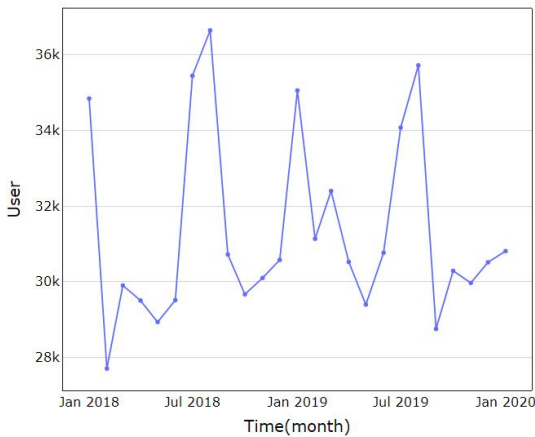


Fig. 1. User by Month

요일별 시간대별 도서관 이용자의 특성을 살펴보면 Fig. 2와 같이 나타나며, 주말과 평일의 경우 이용자 행

태가 각각 다르게 나타남을 알 수 있다. 월요일의 경우 대다수 도서관 휴관일로 인해 대출 수요량이 낮게 나타나며, 휴관으로 인한 수요 전환이 화요일에 집중되는 것으로 판단된다. 또한, 주말의 경우 자료실 마감 시간 1시간 전인 16시에 이용자 증가 폭이 높게 나타나며, 평일도 마감 시간 1시간 전인 17시에 이용자 수가 높지만, 주말과 비교했을 때, 증가 폭은 높지 않은 것으로 나타났다.

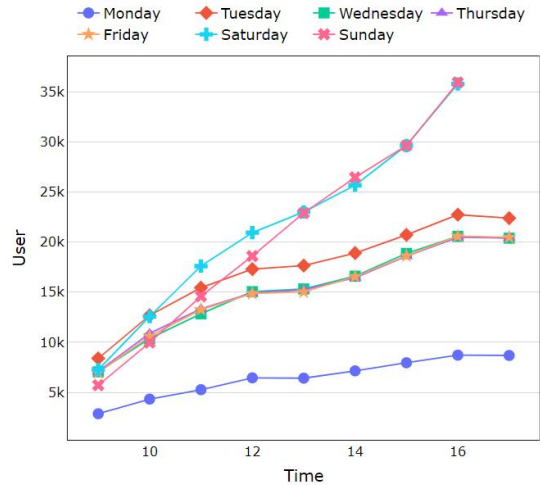


Fig. 2. Average User by Day and Time

3.3.2 시간대별 강우 및 폭염 특성

공공도서관 운영시간 내 시간대별 강우의 특성을 살펴보면 Fig. 3과 같이 나타나며, 평균 36회의 강우가 발생하였으며, 도서관 운영 시작 시각인 9시에 61회로 가장 많이 발생하였고, 17시에 가장 적게 발생하였다.

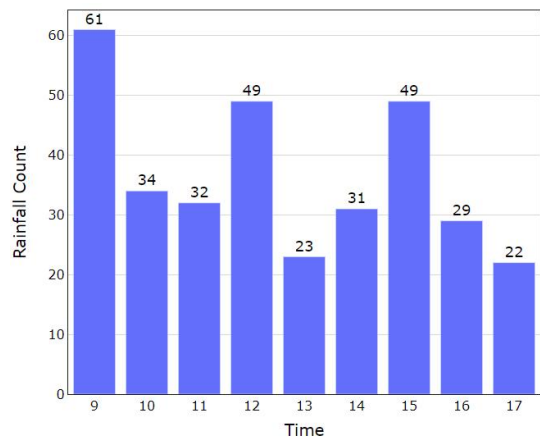


Fig. 3. Rainfall Count by Time

1) 기온뿐만 아니라 습도와 일사량의 영향이 반영된 온열지수(WBGT)를 기반으로 개발된 온도로 대상과 환경에 따라 차별화된 폭염 예방정보를 제공하는 맞춤형 지수

공공도서관 운영시간 내 시간대별 폭염의 특성을 살펴 보면 Fig. 4와 같이 나타나며, 평균 29회의 폭염이 발생하였으며, 오후 시간대인 13시에서 16시에 폭염이 집중적으로 발생하였다.

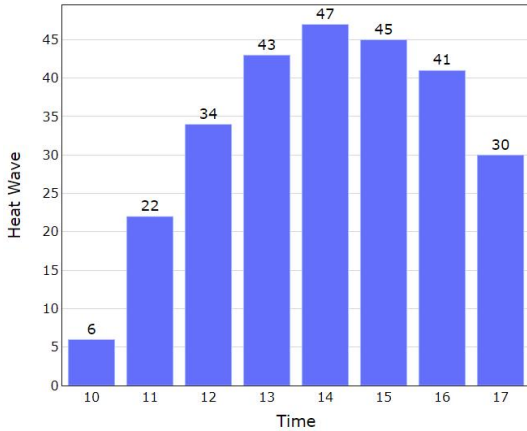


Fig. 4. Heat Wave Count per Time

4. 기상으로 인한 도서 대출 수요 감소

4.1 기상에 의한 도서 대출 수요 감소 정의

본 연구에서는 기상 변화에 따른 공공도서관 대출 수요의 변화를 정량적으로 분석하고, 이에 영향을 미치는 요인을 파악하는 것을 목적으로 한다. 따라서 기상 변화에 따른 대출 수요 감소분을 추정하는 것이 중요하며, 대출 수요의 감소는 강우와 폭염으로 인한 감소로 정의하며, 세부적으로는 도서 대출권 수, 이용자 수, 1인당 도서 대출권 수의 지표를 활용하여 기상에 의한 도서 대출 수요 감소량을 추정하고자 한다.

4.2 기상에 의한 대출 수요 감소 추정 절차

도서 대출 수요 감소를 정량적으로 추정하기 위해서는 시공간적으로 정상시의 도서 대출 수요의 기준값이 필요하며, 도서관 대출 수요에 영향을 미칠 것으로 판단되는 외생변수인 계절성과 요일을 고려하여 4가지 유형으로 데이터를 분류하였다.

앞서 분석된 월별, 요일별, 시간대별 도서 대출 수요 특성에서 학생들의 방학 동안 도서 대출 수요가 높게 나타나 이를 Peak 기간으로 분류하고, 그 외 기간을 Non-Peak로 구분하였다. 또한, 주말과 평일의 경우 도서 대출 수요 패턴이 다르게 나타나 추가로 주말과 평일

을 구분하여, 최종적으로 Table 3와 같이 4가지 유형으로 나누어 데이터 세트를 구축하였다.

Table 3. Types of Borrow Demand

Type	Description
Peak_Weekday	Weekday Borrow Demand in Peak
Peak_Weekend	Weekend Borrow Demand in Peak
Non-peak_Weekday	Weekday Borrow Demand Non-Peak
Non-peak_Weekend	Weekend Borrow Demand Non-Peak

구분된 데이터별로 시공간적으로 정상시의 대출 수요를 파악하기 위해 도서관 시간대별 대출 수요를 산출하고 IQR방식²⁾으로 이상치를 제거한 후 극단치에 영향이 적은 중위값(Median)으로 계산하여 공공도서관별 정상시의 시간대별 공공도서관 대출 수요를 추정하였으며, 세부적인 예시는 Fig. 5와 같이 나타난다. 이를 시공간적 기준 수요(\tilde{D}_{t_n})로 정의하였다.

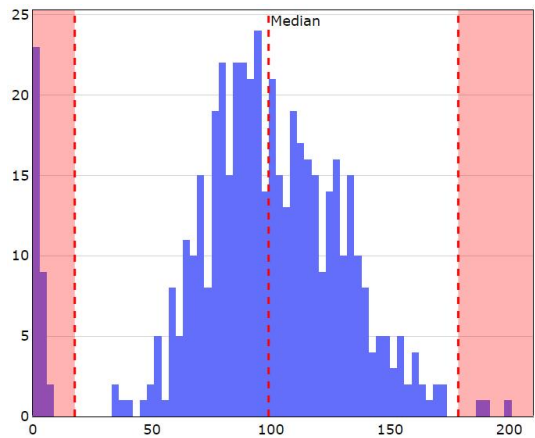


Fig. 5. Calculation of Regular Demand

앞서 계산된 시공간적 기준 수요(\tilde{D}_{t_n})와 공공도서관 시간대별 도서 대출 수요(D_{d,t_n})를 비교하여 감소가 있는 시간대를 확인하였으며, 이를 기상청 기상 데이터의 강우와 폭염이 발생한 구간과의 결합으로 기상으로 인한 대출 수요 감소가 존재하는 구간을 Table

2) 3사분위수(Q3)와 1사분위수(Q1)의 차이로 $Q1-1.5*IQR$ 보다 작고 $Q3+1.5*IQR$ 보다 큰 값을 이상치로 탐지한다.

4와 같이 추정하였다.

Table 4. Estimation of Demand Decrease

Date	Time	Library		
		A	B	...
d_1	t_1	$D_{d_1 t_1 A} - \tilde{D}_{t_1 A}$	$D_{d_1 t_1 B} - \tilde{D}_{t_1 B}$...
d_1	t_2	$D_{d_1 t_2 A} - \tilde{D}_{t_2 A}$	$D_{d_1 t_2 B} - \tilde{D}_{t_2 B}$...
d_1	t_3	$D_{d_1 t_3 A} - \tilde{D}_{t_3 A}$	$D_{d_1 t_3 B} - \tilde{D}_{t_3 B}$...
⋮	⋮	⋮	⋮	⋮

☐ : Rainy, Heat Wave Day

4.3 강우에 의한 도서 대출 수요 감소 추정 결과

강우로 인한 공공도서관별 도서 대출 수요 감소 추정 결과는 Table 5와 같이 나타났으며, 평균적으로 시간당 대출권 수 19.2권, 이용자 수는 5명, 1인당 도서 대출 권 수는 3.8권 감소하였다.

도서관별로는 안심도서관, 북부도서관, 중앙도서관이 대출권 수와 이용자 수의 감소가 높게 나타났으며, 이는 도서관 규모에 기인한 것으로 판단된다.

1인당 도서 대출권 수는 서구 영어도서관, 서구 어린이도서관, 달성도서관, 동부 도서관, 태전도서관이 강우로 인한 도서 대출 수요 감소가 높은 것으로 분석되었으며, 서구에 있는 비원도서관과 원고개도서관의 경우 상대적으로 1인당 도서 대출권 수 감소 폭이 낮게 나타났다.

Table 5. Decrease by Rainfall(Library)

Library	Book (㉔)	User (㉕)	Book per User (㉔/㉕)
2.28 Movement for Democracy	6.0	1.6	3.8
Gususan	22.6	5.3	4.3
Dalseong county	24.9	5.6	4.4
2.28 Students	14.8	3.5	4.2
Nambu	20.0	4.7	4.3
Dalseong	13.0	2.8	4.6
Dongbu	20.6	4.5	4.6
Duryu	19.9	5.7	3.5
Bukbu	30.4	8.0	3.8
Seobu	19.2	4.5	4.3
Suseong	28.1	7.8	3.6
Jungang	30.0	8.9	3.4
Seogu Child	11.0	2.1	5.2

Daemyung eoulrim	3.4	1.0	3.4
Daehyeon	13.2	3.2	4.1
Bisan	4.5	1.3	3.5
Biwon	3.2	1.3	2.5
Seogu English	6.3	1.4	4.5
Sincheon	10.7	3.2	3.3
Ansim	43.0	9.9	4.3
Won Gogae	3.3	1.1	3.0
Icheon	10.6	2.5	4.2
Teajeon	17.1	3.7	4.6
Mean	19.2	5.0	3.8

기상청에서는 강수량에 따라 빗방울(0-1mm/h), 약한비(1-3mm/h), 비(3-15mm/h)의 3가지 유형으로 구분하고 있으며, 강수량에 따른 도서 대출 수요 감소량을 살펴보면, 강수량 증가 시 평균 도서 대출권 수와 이용자 수는 Table 6과 같이 나타났으며, 강수량에 따른 차이는 크지 않은 것으로 보인다.

Table 6. Decrease by Rainfall(Rainfall)

Rainfall	Book (㉔)	User (㉕)	Book per User (㉔/㉕)
0-1(mm/h)	18.76	4.88	3.8
1-3(mm/h)	19.27	5.05	3.8
3-15(mm/h)	19.76	5.13	3.9
Mean	19.2	5.0	3.8

4.4 폭염에 의한 도서 대출 수요 감소 추정 결과

폭염으로 인한 공공도서관별 대출권 수, 이용자 수, 1인당 도서 대출권 수 감소는 Table 7과 같이 나타내며, 평균적으로 대출권 수는 19.2권, 이용자 수는 4.8명, 1인당 도서 대출권 수는 4.0권 감소하였다.

강우로 인한 감소와 유사하게 안심도서관, 중앙도서관에서 대출권 수와 이용자 수 감소 폭이 높게 나타내며, 1인당 도서 대출권 수는 달성도서관과 태전도서관에서 감소 폭이 높으며, 서구 비원도서관, 수성도서관, 대명어울림도서관에서는 감소 폭이 낮게 나타났다.

따라서 기상과 폭염에 따라 공공도서관별 도서 대출 수요 감소가 다름을 알 수 있으며, 도서관의 특성과 공간적인 위치에 따라 일부 군집화 특성을 보이는 것으로 판단된다.

Table 7. Decrease by Heat Wave(Library)

Library	Book (A)	User (B)	Book per User (A/B)
2.28 Movement for Democracy	6.7	1.6	4.2
Gususan	26.0	6.4	4.1
Dalseong county	24.7	5.2	4.8
2.28 Students	16.6	3.6	4.6
Nambu	19.2	4.4	4.4
Dalseong	15.2	2.8	5.4
Dongbu	18.8	3.7	5.1
Duryu	14.0	3.9	3.6
Bukbu	28.3	5.8	4.9
Seobu	21.6	5.0	4.3
Suseong	34.3	10.6	3.2
Jungang	30.5	8.9	3.4
Seogu Child	12.1	2.6	4.7
Daemyung eoulrim	3.5	1.1	3.2
Daehyeon	15.6	3.6	4.3
Bisan	5.2	1.5	3.5
Biwon	4.1	1.3	3.2
Seogu English	6.3	1.4	4.5
Sincheon	13.4	3.9	3.4
Ansim	53.1	14.9	3.6
Won Gogae	3.5	1.0	3.5
Icheon	11.8	2.6	4.5
Teajeon	20.5	4.2	4.9
Mean	19.2	4.8	4.0

폭염 이상의 체감 온도일 경우 1℃ 단위로 나누어 대출권 수, 이용자 수, 1인당 도서 대출권 수를 살펴보면 Table 8과 같이 나타나며, 35℃ 이상의 폭염일 경우 대출권 수, 이용자 수 모두 감소함을 알 수 있다.

Table 8. Decrease by Heat Wave(Temperature)

Temperature	Book (A)	User (B)	Book per User (A/B)
33-34(℃)	19.5	4.9	4.0
34-35(℃)	18.5	4.8	3.9
35-36(℃)	21.7	5.3	4.1
Mean	19.2	4.8	4.0

4.5 경우, 폭염에 의한 도서 대출 수요 감소 군집 분석

본 연구에서는 공공도서관별 경우와 폭염으로 인한 도서 대출 수요 감소의 주요한 요인을 파악하기 위해 감소 폭이 민감한 지역과 민감하지 않은 지역을 군집화하였다. 민감한 지역은 상대적으로 기상변화에 따른 도서 대출 감소 폭이 높은 지역을 의미하며, 민감하지 않은 지역은 감소 폭이 상대적으로 낮은 지역을 의미한다.

데이터 세트의 구조는 Table 9와 같으며, 도서관별 규모의 영향을 배제하기 위해 도서관별 전체 도서 대비 감소한 도서 대출 비율을 산출하고, 이를 정규화하여 적용하였다. 그리고 머신러닝에서 군집 분석으로 주로 활용되는 K-Means³⁾ 알고리즘을 적용하여 기상변화로 인해 공공도서관 대출 수요 감소가 민감한 지역과 민감하지 않은 지역으로 이분화하여 군집화하였다.

Table 9. Dataset for Clustering(example)

Library	Book	User	Book per User
A	0.27	0.33	0.28
B	0.26	0.84	0.34
C	0.31	0.14	0.38
⋮	⋮	⋮	⋮

강우에 따른 대출권 수, 이용자 수, 1인당 대출권 수의 비율 자료를 활용하여 도서관별 군집 분석을 수행한 결과 Fig. 6과 같이 나타나며, 군집의 품질을 정량적으로 평가하는 실루엣 계수(Silhouette Coefficient)를 확인한 결과, 군집이 2개일 때 군집화의 품질이 높으며, 데이터 또한 균등하게 분포하는 것을 알 수 있다.

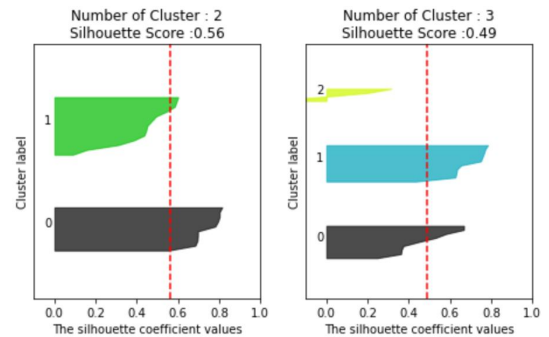


Fig. 6. Rain Component Silhouette Score by Cluster

3) K-Means : 주어진 데이터를 k개의 클러스터로 데이터들의 평균 거리를 기준으로 군집화하는 알고리즘

강우로 인한 군집 분석을 수행한 결과 Table 10과 같으며, 0번 군집에는 10개 도서관, 1번 군집에는 13개의 도서관이 포함되었다. 군집의 특성을 파악하기 위해 군집 별로 대출권 수, 이용자 수, 1인당 도서 대출권 수의 평균 비율을 산출한 결과 0번 군집은 1번 군집에 비해 강우로 인한 도서 대출 감소 폭이 상대적으로 낮았으며, 1번 군집의 경우 강우로 인한 도서 대출 감소 폭이 높음을 알 수 있다.

Table 10. Rain Component Mean by Cluster

Label	0(Insenstive)	1(Sensitive)
Library Count	10	13
Book	0.31	0.40
User	0.30	0.44
Book per User	0.20	0.32

폭염에 따른 대출권 수, 이용자 수, 1인당 도서 대출권 수의 비율 자료를 활용하여 도서관별 군집 분석을 수행한 결과 Fig. 7과 같이 나타나며, 군집의 품질은 앞서 수행한 실루엣 계수로 확인한 결과, 군집이 2개일 때 군집화의 품질이 높으며, 데이터 또한 균등하게 분포하는 것을 알 수 있다.

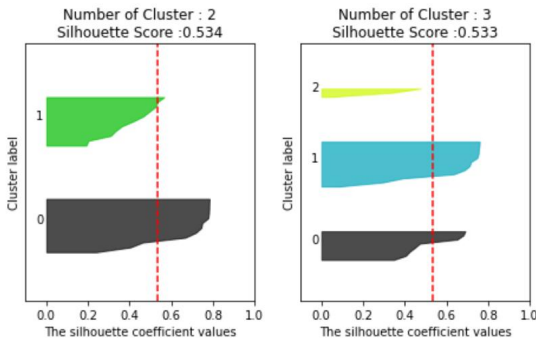


Fig. 7. Heat Component Silhouette Scroe by Cluster

폭염으로 인한 군집 분석을 수행한 결과 Table 11과 같이 나타나며, 0번 군집에는 12개 도서관, 1번 군집에는 11개의 도서관이 포함되었다. 군집의 특성을 파악하기 위해 군집 별로 대출권 수, 이용자 수, 1인당 도서 대출권 수의 평균 비율을 산출한 결과, 0번 군집은 1번 군집에 비해 폭염으로 인한 감소 폭이 상대적으로 낮았으며, 1번 군집이 폭염에 민감하게 감소 폭이 높은 것으로 분석되었다.

Table 11. Heat Component Mean by Cluster

Label	0(Insenstive)	1(Sensitive)
Library Count	12	11
Book	0.34	0.45
User	0.30	0.50
Book per User	0.18	0.31

4.5 도서 대출 수요 감소 군집별 요인 중요도

기상 변화에 따른 공공도서관 도서 대출 수요에 영향을 미치는 요인들을 분석하기 위해 감소한 대출권 수, 이용자 수, 1인당 도서 대출권 수의 비율로 산정된 군집 라벨을 종속변수로 설정하고, 공공도서관의 내적 요인과 외적 요인, 인구·통계학적 변수들을 독립변수로 설정하여 대표적인 머신러닝 알고리즘인 랜덤 포레스트로 분류 예측모형을 구축하고, 요인별 특성 중요도(Feature Importance)⁴⁾를 도출하고자 한다.

모형구축에 활용된 변수는 총 23개로 세부적 사항은 Table 12와 같이 나타나며, 변수 구성은 선행연구에서 공공도서관 이용 현황에 영향을 미치는 요인들을 참조하여 구성하였다. 또한, 앞서 수행된 선행연구에서 검토하지 않은 요인들을 추가로 반영하여 분석을 수행하였으며, 내적 요인으로는 도서의 수, 도서관 좌석 수, 도서관 접근 경사도 등이 있으며, 외적 요인으로는 대표적인 빅데이터인 통신사 유동인구 데이터와 카드 매출실적 자료를 활용하였다. 그 외에도 공공도서관을 학업적인 목적으로 대체할 수 있는 시설인 카페, 서점, 독서실과의 거리, 교통 접근성, 교육시설 등의 변수를 반영하여 모형을 고도화하였다.

교통 접근성의 경우 도서관 반경 800M 이내에 존재하는 버스 정류장과 지하철역의 개수, 도서관까지의 평균 도보거리를 산정하여 변수에 포함하였으며, 도보거리의 경우 거리를 산정하는 파이썬 라이브러리인 openrouteservice를 활용하여 최단 거리를 계산하였다. 또한, 배후지역의 교육시설 및 유동인구의 수, 업종별 평균 소비 자료는 통상적으로 도보로 걸을 수 있는 10분 반경의 범위를 800M를 영향권으로 설정하여 데이터를 수집하여 변수로 반영하였다.

4) 특성 중요도(Feature Importance) : 의사 결정 트리는 노드의 불순도를 가장 크게 감소시키는 변수로 마디를 형성하는데, 이를 측정하는 지니계수의 평균 감소량(mean decrease of Gini)으로 특성 중요도를 판단하며, 지니평균 감소량이 많을수록 중요도가 높은 변수가 된다.

Table 12. Rain / Heat Variables in Random Forest Model

Category	Variable	Description	Format	
D e p e n d e n t	Cluster Label	Rain / Heat Cluster Label	Types of decrease in loan demand due to weather	Categorical
	Library	Number of books	The number of book of Public Library	Continuous
Number of seats		The number of seat of Public Library	Continuous	
Slop		Library Access slope	Continuous	
Area		Total ground area of Public Library	Continuous	
Consumption	Average sales for cafe industries	Average sales of all industries	Continuous	
	Average sales for book industries	Average sales of books	Continuous	
	Average sales for reading room industries	Average sales of the reading room industry	Continuous	
Alternative facility	Distance to cafe	Average distance to cafes	Continuous	
	Distance to bookstore	Average distance to bookstores	Continuous	
	Distance to reading room	Average distance to reading rooms	Continuous	
	Distance to small library	Average distance to small libraries	Continuous	
Traffic	Distance to busstop	Mean of Bus Stop Distance	Continuous	
	Distance to subway	Mean of Subway Station Distance	Continuous	
	Number of busstop	Number of Bus Stop	Continuous	
	Number of subway stations	Number of Subway Station	Continuous	
Education	Number of elementary schools	Number of Elementary School	Continuous	
	Number of middle schools	Number of Middle School	Continuous	
	Number of high schools	Number of High School	Continuous	
Population	10's male /female floating population	Sum of 10s Male/female floating population	Continuous	
	60's over male /female floating population	Sum of 60s up male/female floating population	Continuous	

구축된 랜덤 포레스트로 모형으로 강우에 따른 도서관 도서 대출 수요에 영향을 미치는 요인들을 특성 중요도로 산출한 결과 Fig. 8과 같이 나타났으며, 공공도서관의 내적 요인인 도서 수, 도서관 면적, 좌석 수의 요인이 가장 중요한 요인으로 나타났다. 또한, 외적 요인으로는 도서관 접근 경사로, 주변 카페 평균 매출, 주변 카페/서점과의 거리, 버스 정류장 수가 높은 것으로 분석되었다.

인구통계학적으로는 차량 운전이 불가한 10대 유동인구 변수의 중요도가 높게 나타나며, 30/40대 여성도 강우로 인한 도서 대출 수요 감소가 민감하게 나타났다. 그의 교육시설 수(초/중/고등학교)는 강우로 인한 도서 대출 수요 감소에 영향이 낮은 것으로 분석되었다.

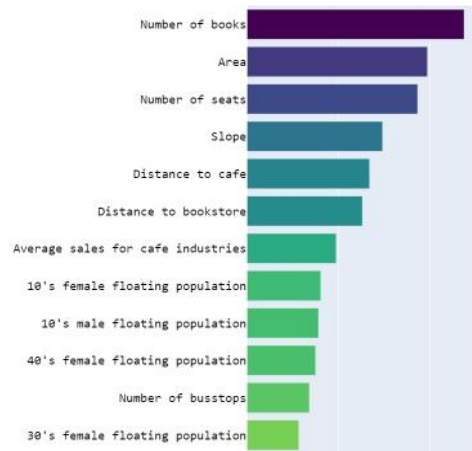


Fig. 8. Feature Importance of Rain Cluster

폭염도 강우와 유사하게 공공도서관 내적 요인인 좌석 수, 도서의 수, 면적이 가장 중요한 변수로 Fig. 9와 같이 나타나며, 공공도서관과 유사한 학업 공간을 제공하는 주변 카페와 독서실 관련 변수가 중요한 것을 알 수 있다. 인구통계학적으로는 강우와 유사하게 10/20/30대 여성들이 폭염으로 인한 공공도서관 도서 대출 수요 감소에 민감하게 반응하는 것으로 분석되었으며, 교통시설 측면에서는 버스 정류장의 수의 영향이 일부 존재하는 것으로 나타났다. 또한, 기상과 폭염의 공통점으로는 공공도서관의 규모가 크고, 공공도서관 인근의 카페와의 거리가 중요한 요인으로 분석되었고, 인구통계학적으로는 10대와 20/30대 여성 계층들이 기상 변화에 민감하게 반응하는 것으로 나타났다. 하지만 강우와 폭염의 차이점은 기상의 경우 공공도서관 접근 경사도가 주요한 변수로 나타난 반면, 폭염의 경우 영향이 낮은 것으로 분석되었다.

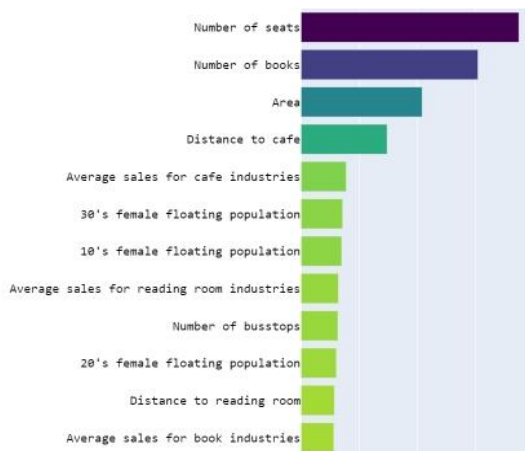


Fig. 9. Feature Importance of Heat Cluster

5. 결론

5.1 분석결과 요약

본 연구는 대구시 공공도서관 도서 대출 이력 데이터와 기상청 종관기상관측 자료를 수집하여, 기상으로 인한 도서 대출 수요 감소를 정량적으로 산출하고, 수요 감소에 영향을 미치는 주요한 요인 도출을 목적으로 하였다.

이를 시공간적으로 분석하기 위해 공공도서관별 평상시의 도서 대출 수요를 IQR 기법으로 이상치를 제거한 후 중위값을 기준으로 산출하였으며, 산출된 기준값과 도서관별 시간대에 따른 도서 대출 수요의 차이를 계산하여 강우와 폭염으로 인한 도서 대출 감소량을 추정하였다. 추정된 감소량을 기반으로 기상으로 도서 대출 감소가 상대적 높은 지역과 그렇지 않은 지역으로 분류하고, 이를 군집화하여 종속변수로 설정하였다. 또한, 앞서 선행연구에서 검토된 공공도서관 내적 요인과 외적 요인, 통신사 유동인구, 카드 매출 등의 빅데이터와 결합하여 기상으로 인한 공공도서관 대출 수요 요인분석을 수행하였다.

본 연구에서의 수행된 주요 분석 결과는 다음과 같다.

첫째, 강우와 폭염으로 인한 도서 대출 수요 감소가 높은 도서관은 안심, 북부, 수성, 중앙도서관으로 나타났으며, 이들은 지역의 대표적인 거점 도서관이다. 그리고 도서관의 특성과 공간적인 위치에 따라 일부 군집화 특성을 보이는 것으로 분석되었다.

둘째, 기상의 경우 강우량에 따른 감소 차이는 크게 없었으며, 폭염은 35°C 이상일 경우 대출권 수, 이용자

수 모두 감소하는 것으로 분석되었다.

셋째, 기상 변화에 의한 공공도서관의 대출 수요 감소를 군집화한 결과, 상대적으로 민감한 지역(도서 대출 수요가 감소가 높은 지역)과 민감하지 않은 지역(도서 대출 수요 감소가 낮은 지역)으로 분류되었으며, 분류된 공공도서관이 강우와 폭염 모두 유사한 것으로 나타났다.

넷째, 머신러닝으로 기상 변화에 따른 공공도서관 대출 수요 차이의 영향 요인을 분석한 결과, 공통적으로 도서 수, 면적, 좌석 수와 같은 내적 요인과 주변 카페와의 거리, 카페 매출 등의 외적 요인 등이 중요한 요인으로 분석되었다.

마지막으로 인구통계학적으로는 10대와 20/30 여성일 경우 기상으로 인한 도서 대출 수요가 민감하게 반응하는 것으로 나타나며, 강우의 경우 공공도서관의 접근 경사도가 도서 대출 수요 감소의 주요한 요인으로 도출되었다.

5.2 연구 시사점 및 한계

본 연구의 시사점은 다음과 같다.

첫째, 기상에 의한 공공도서관 대출 수요 감소를 정량적으로 도출한 점과 수요 감소에 영향을 주는 요인들을 다양하게 살펴본 점이 선행연구와의 차별성이라 할 수 있으며, 이러한 분석 결과가 특정 시점 공공도서관 이용 활성화 정책에 활용될 수 있을 것으로 판단된다.

둘째, 앞서 선행연구에서 검토된 요인 외에도 공공도서관 접근 경사로, 통신사 유동인구, 카드 매출 데이터 등의 빅데이터를 결합하여 주요한 요인들을 도출하였으며, 이를 기반으로 특정 시점 공공도서관 이용자의 불편함과 만족도 개선에 이용될 수 있을 것으로 기대한다.

셋째, 인구통계학적으로 10대, 20/30대 여성 계층에서 기상으로 인한 공공도서관 도서 대출 수요 감소에 민감한 것으로 분석되었으며, 해당 계층을 대상으로 하는 마케팅 및 이용 활성화 정책을 개발한다면 공공도서관 이용 활성화에 이바지할 것으로 보인다.

마지막으로 교육, 교통 관련 요인들은 상대적으로 공공도서관 내적 요인과 소비, 인구·통계학적, 학업 대체 시설보다 영향 정도가 낮은 것으로 분석되었으며, 추후 기상과 공공도서관 대출 수요의 후속 연구에 있어 기초 자료로 활용될 수 있을 것으로 기대한다.

본 연구의 한계점은 다음과 같다.

첫째, 본 연구에서는 첨두와 비첨두 시점과 주말, 평일 그리고 코로나19로 인한 영향을 제거하여 분석을 수행하였으나, 그 외 고려되지 않은 외생변수 요인이 존재

할 것으로 판단되며, 향후 외적 요인들을 제거하여 기상으로 인한 도서 대출 수요 감소를 추정한다면 더욱 정확한 결과를 도출할 수 있을 것으로 판단된다.

둘째, 기상의 경우 공간적으로 다르고 시간에 지남에 따라 지속해서 변화됨을 고려하면, 대구시 이외에도 강원도, 제주도, 수도권 등의 다양한 분석 결과와 종합하여 비교 분석한다면 의미 있는 연구 결과가 도출될 것으로 기대한다.

셋째, 공공도서관 도서 대출 이력을 기반으로 기상변화에 따른 수요 감소를 추정하였는데, 공공도서관의 경우 도서 대출 이외에도 문화행사 참여, 평생교육, 열람실 이용 등의 다양한 목적의 방문객들이 존재하는바 추후 이용자 방문 이력 자료와 결합한다면 다양한 측면에서 분석이 가능할 것으로 판단된다.

마지막으로 최근 지구 온난화로 인한 기상변화와 신종 감염병 등 다양한 사회적 이슈들이 발생하고 있으며, 기상변화와 관련된 연구는 증가할 것으로 판단된다. 따라서 기상과 공공도서관과의 연구가 활성화되어 공공도서관 활성화에 이바지하기를 기대한다.

REFERENCES

- [1] Committee on Library and Information Policy. (2008). *The 1st comprehensive library advancement plan 2009-2013*. Seoul: Presidential Committee on Library and Information Policy
- [2] Committee on Library and Information Policy. (2014). *The 2nd comprehensive library advancement plan 2014-2018*. Seoul: Presidential Committee on Library and Information Policy
- [3] Committee on Library and Information Policy. (2019). *The 3rd comprehensive library advancement plan 2019-2023*. Seoul: Presidential Committee on Library and Information Policy
- [4] J. K. Park & I. Kim. (2016). An Analysis of Geographical Accessibility to Public Libraries in Metropolitan Area and Related Implications: The Case of Busan. *The Korean Journal of Local Government Studies*, 20(3), 191-221.
DOI : 10.20484/klog.20.3.10
- [5] G. H. Jeon & S. Y. Kwon. (2018). A Study on the Influence of Economic Factors on Library Use. *Journal of the Korea Convergence Society*, 9(11), 299-306.
DOI : 10.15207/JKCS.2018.9.11.299
- [6] H. J. Park, S. M. Jun & J. C. Jung. (2018). The impact of public facilities on the perception of parents visiting public libraries. *Journal of Korea Planning Association*, 53(4), 23-38.
DOI : 10.17208/jkpa.2018.08.53.4.23
- [7] K. J. Lee. (2020). The Influence Factors on the Numbers of Visitors and Reference Room Users of Public Libraries: Based on the National Libraries Statistical Data 2018. *Journal of the Korean Library and Information Science*, 54(2), 105-125.
DOI : 10.4275/KSLIS.2020.54.2.105
- [8] T. Y. Kim, J. Y. Baek & H. J. Oh. (2018). An Analysis of Library User and Circulation Status based on Bigdata Logs A Case Study of National Library of Korea. Sejong. *Journal of Korean Library and Information Science Society*, 49(2), 357-388.
DOI : 10.16981/kliss.49.201806.357
- [9] Y. S. Chung. (2012). Assessment of non-recurrent congestion caused by precipitation using archived weather and traffic flow data. *Transport Policy*, 19(1), 167-173
DOI : 10.1016/j.tranpol.2011.10.001
- [10] K. W. Shin & K. C. Choi. (2014). Analyzing the Relationship Between Precipitation and Transit Ridership Through a Seemingly Unrelated Regression Mode. *Journal of Korean Society of Transportation*, 32(2), 83-92
DOI : 10.7470/jkst.2014.32.2.083
- [11] M. S. Won, S. H. Cheon, S. I. Shin & S. Y. Lee. (2019). Analysis of Public Transport Ridership during a Heavy Snowfall in Seoul. *KSCE Journal of Civil and Environmental Engineering Research*, 39(6), 859-867
DOI : 10.12652/Ksce.2019.39.6.0859
- [12] J. H. Lee & H. Y. Jung. (2018). The Impact of Weather Conditions on Transit Ridership Using Quantile Regression Analysis. *Journal of Korea Planning Association* 53(4), 95-106
DOI : 10.17208/jkpa.2018.08.53.4.95
- [13] S. S. Lee. (2011). Applications of Geographic Information Systems in Public Library Marketing. *Journal of the Korean society for information management*, 28(3), 179-195.
DOI : 10.3743/KOSIM.2011.28.3.179
- [14] K. Y. Park & S. B. Lee. (2012). A Study on the Effect of Adverse Weather Conditions on Public Transportation Mode Choice. *KSCE Journal of Civil and Environmental Engineering Research*, 32(1D), 23-31.
DOI : 10.12652/Ksce.2012.32.1D.023
- [15] M. K. Oh, K. R. Kim, W. O. Jeong & K. W. Kim. (2021). Analysis of Borrows Demand for Books in Public Libraries Considering Cultural Characteristics. *Journal of Digital Convergence*, 19(3), 55-64.
DOI : 10.14400/JDC.2021.19.3.055

오 민 기(Min-Ki Oh)

[학생회원]



- 2016년 3월 ~ 현재 : 경북대학교 문헌정보학과
- 2020년 7월 ~ 12월 : 대구디지털산업진흥원 인턴
- 2020년 6월 ~ 8월 : 대구디지털산업진흥원 인턴
- 관심분야 : 빅데이터, 머신러닝, 텍스트 마이닝, NLP

트 마이닝, NLP

· E-Mail : dmg04023@gmail.com

장 원 준(Won-Jun Chang)

[학생회원]



- 2021년 2월 : 영남대학교 통계학(학사)
- 2020년 11월 ~ 현재 : 대구디지털산업진흥원 빅데이터활용센터 전임연구원
- 2022년 2월 ~ 현재 : 경북대학교 데이터사이언스 대학원
- 관심분야 : 빅데이터, 머신러닝, 텍스트 마이닝, 데이터 시각화

마이닝, 데이터 시각화

· E-Mail : bd1jun@dip.or.kr

김 건 옥(Keun-Wook Kim)

[강사]



- 2009년 2월 : 영남대학교 도시공학(공학사)
- 2011년 8월 : 아주대학교 교통공학(공학석사)
- 2019년 7월 ~ 현재 : 대구디지털산업진흥원 빅데이터활용센터 센터장
- 관심분야 : 도시데이터분석, 빅데이터, 인공지능, 텍스트마이닝

인공지능, 텍스트마이닝

· E-Mail : aut7767@dip.or.kr

신 세 영(Se-Young Shin)

[학생회원]



- 2017년 3월 ~ 현재 : 경북대학교 글로벌소프트웨어융합학과
- 2021년 7월 ~ 8월 : 대구디지털산업진흥원 인턴
- 관심분야 : 빅데이터, 머신러닝
- E-Mail : tlstpdud98@gmail.com

이 진 명(Jin-Myeong Lee)

[학생회원]



- 2016년 3월 ~ 현재 : 경북대학교 물리학과
- 2021년 6월 ~ 8월 : 대구디지털산업진흥원 인턴
- 관심분야 : 빅데이터, 머신러닝
- E-Mail : apzhs@naver.com