



머신러닝을 이용한 소화용수시설 근처 불법주정차 발생 예측

김여림(경희대학교 석사 과정, erin9906@khu.ac.kr), 이용채(경희대학교 학사 과정, yong03111@khu.ac.kr)

I. 연구 개요

- 현대 사회에서 대도시권의 불법주정차는 심각한 사회문제로 인식되고 있으며, 특히 **응급상황에서 불법주정차된 차량은 큰 문제로 이어질 수 있음.**
 - 소방차 전용구역 및 소화용수시설 주변의 불법주정차는 긴급차량의 진입 및 통행, 시설 이용 등을 방해하여, 대표적인 골든타임 지체 원인임(한학 외, 2022).
 - 의정부 대봉 그린 아파트 화재와 제천 스포츠 센터 화재의 경우, 불법주정차 차량으로 인해 소방차 진입이 늦어져 더 큰 피해를 입은 바 있음(최유정 외, 2022).
- 소방차량의 효율적인 동선 설정과 소화용수시설 이용을 위해 **실시간으로 소화용수시설 근처의 불법주정차 발생 예측**이 필요하지만, 관련 연구가 부족한 실정임.
 - 불법주정차 관련 연구는 설문조사 등의 전통적인 기법을 통한 연구부터(정현영 외, 1999), 공간회귀분석을 통한 연구까지 (박준상·이수기, 2022) 불법주정차 선택 원인이나 유발 요인을 탐색하고자 하는 연구가 주를 이루었기 때문에 예측을 시도한 연구는 매우 부족한 실정임.
- 따라서 **머신러닝 알고리즘을 활용하여 소화용수시설 근처 불법주정차의 발생 여부를 예측**하고자 함
 - 연구지역은 강남구로, 서울의 중심지로서 교통량이 매우 많고 상업 및 주거 지역이 혼재하는 지역이기 때문에 머신러닝을 적용하기 적합한 지역일 것으로 판단하여 선정함.
 - 최근 머신러닝을 이용해 예측 모델을 구축하는 연구가 부동산(김정희, 2021; 배성완·유정석, 2018), 교통(이호준·이수기, 2021; 남명우 외, 2021) 등의 다양한 분야에서 이루어지고 있음.
 - 불법주정차 문제에 있어서도 머신러닝 예측 모델을 활용한다면 화재 응급대응에 있어 유의미한 효과를 가져올 수 있을 것으로 기대됨.

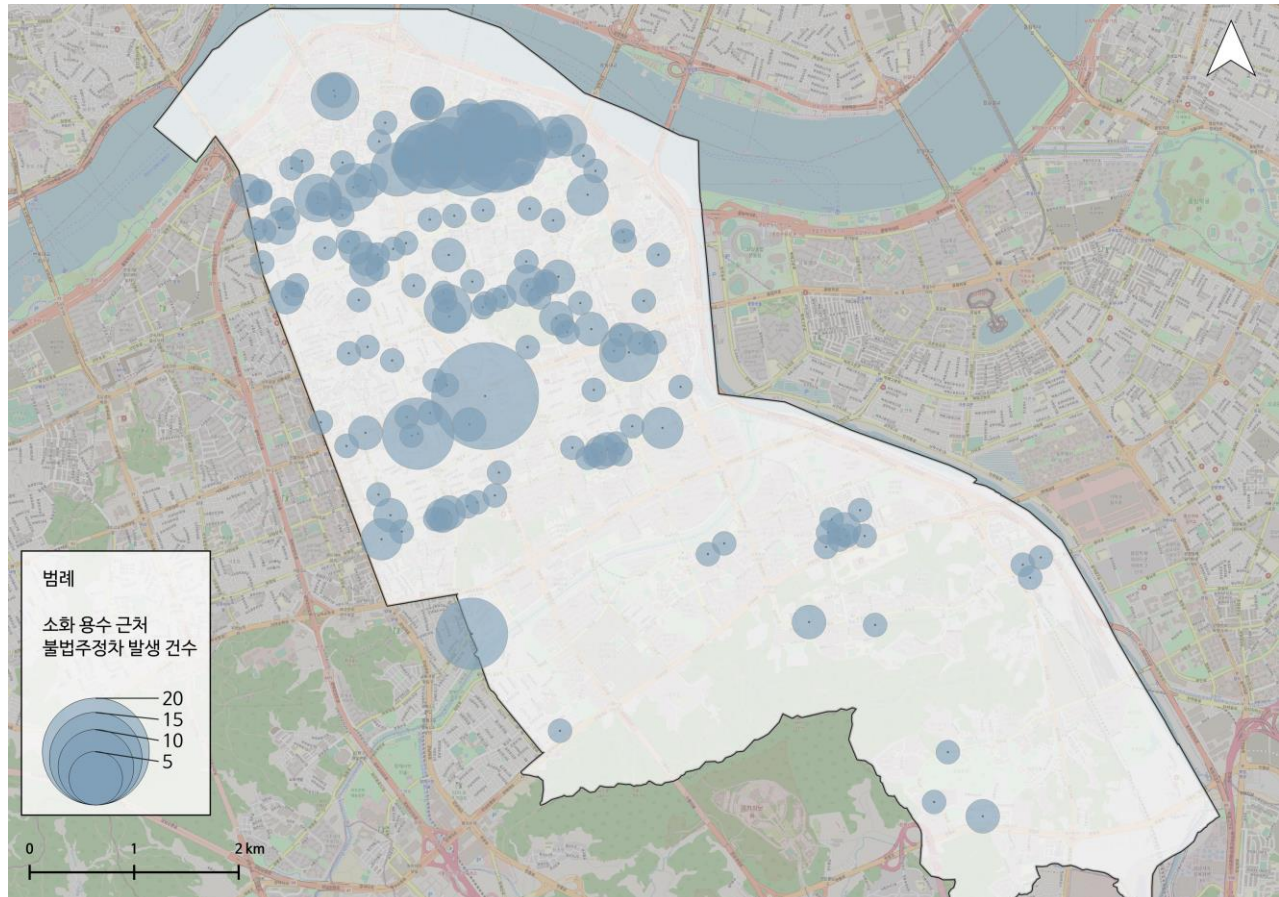


그림 1. 소화용수 근처 불법주정차 발생 건수

II. 사용 변수

- 독립변수는 **공간적 변수**와 **시계열적 변수**로 구분 지을 수 있음.
 - 공간적 변수 중 집계구별 변수의 경우, 공간자기상관을 고려하고자 공간 가중 행렬을 곱하여 사용함.
 - 시계열적 변수 중 생활인구 및 기상요소는 시간대별로 집계하였으며, 시간대는 심야/새벽, 아침, 오전, 점심, 오후, 저녁, 밤으로 구분함.
 - 집계구별 건물용도별 면적의 경우, GIS건물통합정보 데이터의 세부 용도를 사용하여 아파트/다세대, 단독/다가구, 교육시설, 상업시설, 업무시설, 의료/복지시설, 음식점으로 재분류함.
 - 상관관계 분석과 VIF지수를 통해 다중공선성 문제를 확인하였으며, 집계구별 의료/복지시설, 음식점 연면적 변수를 제거하였음.
- 종속변수는 **소화용수별 시간대별 불법주정차 발생 유무**임.

표 1. 연구 변수 및 데이터 출처

연구 변수	구분	자료형	데이터 출처
집계구별 건물용도별 면적 (아파트/다세대, 단독/다가구, 교육시설, 상업시설, 업무시설)	공간적 독립 변수	연속형(정규화)	국가공간정보포털
집계구별 주차장 면수		연속형(정규화)	공공데이터포털
최근린 주차장 정보(거리, 면수)		연속형(정규화)	공공데이터포털
집계구별 생활인구 (0-19세, 20-64세, 65세 이상)	시계열적 독립 변수 (2019,2021)	연속형(정규화)	서울열린데이터광장
기상 요소(동별 시간대별 기온 및 강수량)		연속형(정규화)	기상자료개방포털
시간대, 요일(주중/주말), 월		범주형(더미변수)	서울빅데이터캠퍼스
2019년 소화용수 근처 불법주정차 발생 유무	종속변수	범주형(이진형)	서울빅데이터캠퍼스
2021년 소화용수 근처 불법주정차 발생 유무	종속변수	범주형(이진형)	공공데이터포털

III. 연구 방법

- 여러 머신러닝 분류 알고리즘을 학습하여 **소화용수별 시간대별 불법주정차 발생 유무**를 예측하고 비교하여 가장 높은 정확도를 가진 모델을 선정함.
 - 학습은 2019년의 변수를 기준으로 진행하고, 정확도는 2021년도의 변수를 기준으로 측정함.
 - 2021년의 불법주정차 발생 여부를 예측함으로써, 개발한 모델의 적용 가능성을 확인해보고자 함.
- 학습 데이터셋의 불균형을 해소하고자 종속변수가 1인 데이터를 기준으로 Under sampling을 진행하였으며, 예측 데이터셋의 경우 원데이터의 부정확한 주소 체계로 인한 불확실성을 극복하고자 지오코딩 결과 위치를 가장 가까운 도로로 이동시킴.
- 연구에서 사용한 모델은 다음과 같음.
 - 랜덤포레스트(Random Forest): 앙상블 학습으로, 여러 개의 의사결정 나무를 조합하여 예측 모델을 생성함.
 - 로지스틱회귀(Logistic Regression): 종속 변수가 범주형일 때 사용되며, 입력 변수와 종속 변수 간의 선형 관계를 모델링하여 확률을 예측하는데 사용됨.
 - SVM(Support Vector Machine): 두 클래스 사이에서 가장 큰 마진(margin)을 가지는 초평면을 찾아내는 모형으로(Vapnik, 2000), 본 연구에서는 선형커널과, RBF(Radial Basis Function)커널을 사용함.
 - LDA/QDA(Linear/Quadratic Discriminant Analysis): 판별분석은 각 클래스의 사전 확률 값을 구한 뒤, 새로운 관측값을 속할 확률이 가장 높은 클래스로 분류하며, LDA는 입력 변수의 선형 조합을, QDA는 2차 조합을 사용함.

IV. 분석 결과

- 각 모형으로 학습 및 예측한 결과, 모형별 검증 정확도와 예측 정확도는 [표 2]와 같음.
 - QDA를 제외하고 모든 모형에서 검증 정확도는 높게 나타났음.
 - 예측의 경우에는, 대부분의 모형에서 50% 이하의 정확도를 나타냈으며, **RBF 커널 기반 SVM 모형이 약 53%로 가장 높은 정확도를 나타냈음.**
- 따라서, 본 연구에서는 **RBF 커널 기반 SVM 모형**을 채택하였음.
 - 모형의 분류 결과는 [표 3]에 해당하며, 소화용수별 분류 결과를 지도로 시각화하였음.([그림 2])
 - 예측에 사용된 2,246개 소화용수 중, 모든 경우를 정확히 예측한 소화용수는 1,768개, 모든 경우를 다르게 예측한 소화용수는 279개에 해당함.

표 2. 모형 학습 및 예측 결과

구분	검증 정확도	예측 정확도
Random Forest	1.000	0.5
Logistic	1.000	0.417
SVM(Linear)	1.000	0.455
SVM(RBF)	1.000	0.532
LDA	1.000	0.455
QDA	0.519	0.5

표 3. 모형 분류 결과

	Predicted 0	Predicted 1
Actual 0	2808	395
Actual 1	2602	601

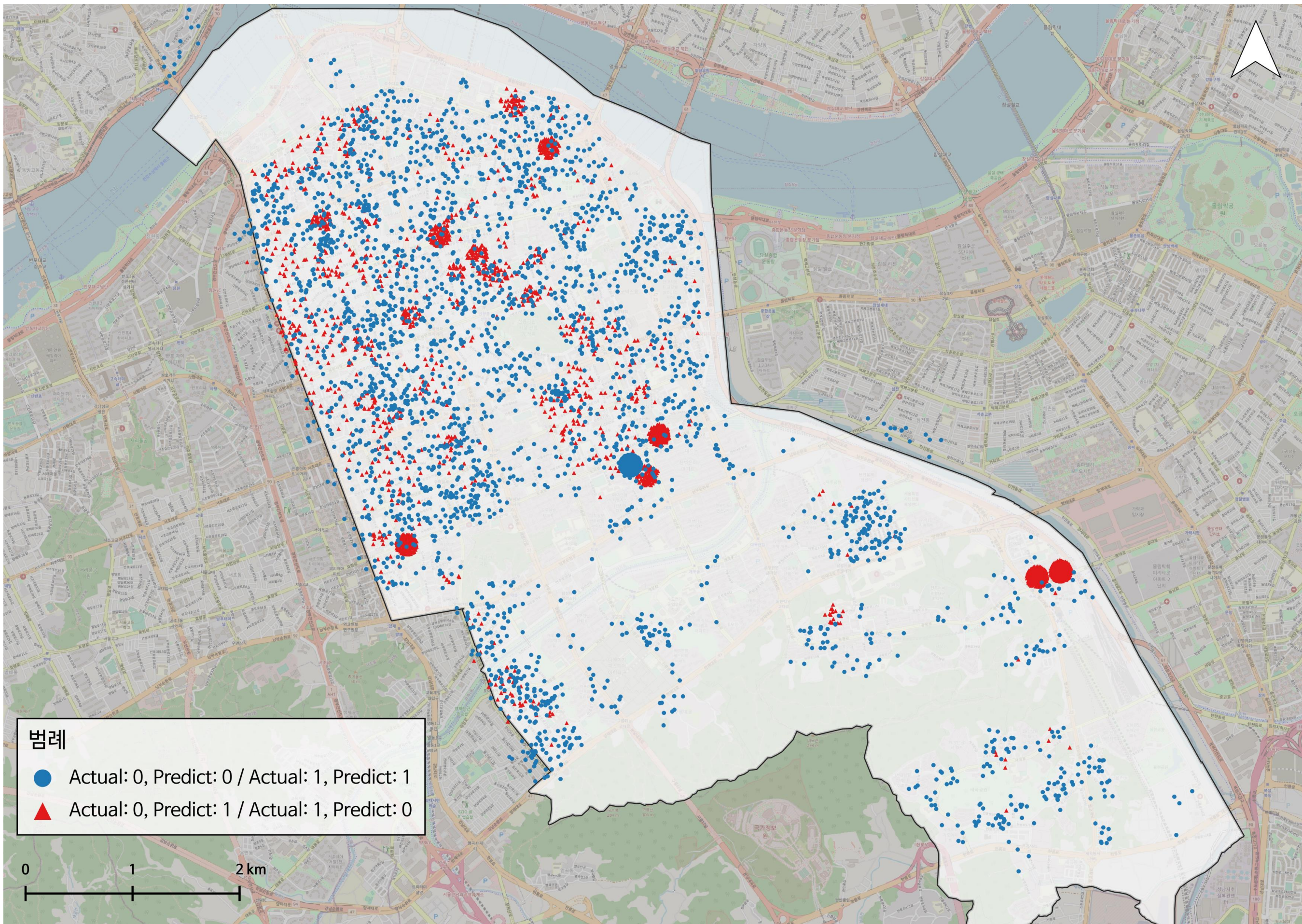


그림 2. 소화용수별 예측 결과

V. 결론

- 본 연구에서는 다양한 머신러닝 알고리즘을 이용해 소화용수시설 근처 불법주정차 발생 여부를 예측하는 모델을 생성한 뒤, 예측 정확도를 기반으로 최적의 모델을 도출함.
 - 랜덤 포레스트, 로지스틱 회귀, SVM, LDA, QDA 등의 알고리즘을 활용하여 불법주정차 발생 데이터를 학습시킨 후, 예측 정확도를 구한 결과 RBF 커널 기반 SVM이 가장 높은 정확도를 보였음.
- 본 연구의 한계점과 추후 보완할 사항은 다음과 같음.
 - 데이터 접근성의 한계로 2019년의 불법주정차 발생 데이터만을 사용했기 때문에 장기간의 데이터로 모델의 학습을 보완할 필요가 있음.
 - 예측에 사용된 2021년 불법주정차 데이터의 위치 불확실성이 예측 정확도를 낮추었을 것으로 예상됨.
 - 파라미터 조정을 통한 모델 최적화는 진행했지만, 최적의 변수 조합을 선정하는 과정은 거치지 못했음.
- 위의 한계점에도 불구하고, 본 연구의 결과는 기존의 불법주정차 연구에서 많이 시도되지 않았던 **불법주정차 발생의 예측**에 초점을 맞추고 그 가능성을 보여주었다는 점에서 의의가 있음.
- 또한 실시간으로 변화하는 변수를 사용함으로써 시계열적 요소까지 고려한 **시공간 예측**을 시도했다는 점에서 의의가 있음.

참고문헌

- Vapnik, V, 2000, The Nature of Statistical Learning Theory 2nd Edition, Springer-Verlag New York, USA, pp. 1-314.
- 김정희. (2021). 머신러닝을 이용한 주택가격 예측과 공간적 군집패턴 분석. 한국사지리학회지, 31(4), 150-165.
- 남명우, 박두서, 장영준, & 이홍철. (2021). 머신러닝을 이용한 교통사고 사상자 수 예측: 서울시 공공데이터를 대상으로. 한국컴퓨터정보학회 학술발표논문집, 29(1), 27-30.
- 박준상, & 이수기. (2022). 불법 주정차에 영향을 미치는 도시 환경 요인 분석: 서울시 스마트 불편신고 민원자료를 중심으로. 지역연구, 38(3), 3-17.
- 배성완, & 유정석. (2018). 머신 러닝 방법과 시계열 분석 모형을 이용한 부동산 가격지수 예측. 주택연구, 26, 107-133. 이창희, 김명수, & 서소민. (2014). 데이터마ining 기법을 활용한 불법주차 영향요인 분석. 한국 ITS 학회 논문지, 13(4), 63-72.
- 이호준, & 이수기. (2021). 보행자 교통사고 심각도 예측을 위한 머신러닝 모형 비교 분석: 보행자 교통사고 자료의 표본 불균형 보정을 중심으로. 대한공간정보학회지, 29(2), 3-15.
- 정현영, 박태래, & 김기영. (1999). 도심지 노상주차 실태와 운전자의 인식. 대한토목학회 학술발표회 논문집, 4, 17-20.
- 최유정, 최수길, & 김시국. (2022). 화재사례 분석을 통한 소방안전관리자용화재대응 훈련 시나리오 개발을 위한 기초연구. Fire Science and Engineering, 36(1), 43-55.
- 한학, 윤홍식, 함태윤, & 홍석범. (2022). System Dynamics-GIS 모델을 이용한 화재 출동 시 강제처분의 효과분석 - 포항시를 중심으로. Crisisonomy, 18(7), 19-31.