

# 머신러닝을 활용한 어린이 스마트 횡단보도 최적입지 선정

- 창원시 사례를 중심으로 -

## Machine Learning based Optimal Location Modeling for Children's Smart Pedestrian Crosswalk: A Case Study of Changwon-si

이수현<sup>1)</sup>, 서용원<sup>2)</sup>, 김세인<sup>3)</sup>, 이재경<sup>4)</sup>, 윤원주<sup>5)</sup>

Lee, Suhyeon<sup>1)</sup> · Suh, Youngwon<sup>2)</sup> · Kim, Sein<sup>3)</sup> · Lee, Jaekyung<sup>4)</sup> · Yun, Wonjoo<sup>5)</sup>

Received February 03, 2022; Received May 06, 2022 / Accepted May 08, 2022

**ABSTRACT:** Road traffic accidents (RTAs) are the leading cause of accidental death among children. RTA reduction is becoming an increasingly important social issue among children. Municipalities aim to resolve this issue by introducing "Smart Pedestrian Crosswalks" that help prevent traffic accidents near children's facilities. Nonetheless such facilities tend to be installed in relatively limited number of areas, such as the school zone. In order for budget allocation to be efficient and policy effects maximized, optimal location selection based on machine learning is needed. In this paper, we employ machine learning models to select the optimal locations for smart pedestrian crosswalks to reduce the RTAs of children. This study develops an optimal location index using variable importance measures. By using k-means clustering method, the authors classified the crosswalks into three types after the optimal location selection. This study has broadened the scope of research in relation to smart crosswalks and traffic safety. Also, the study serves as a unique contribution by integrating policy design decisions based on public and open data.

**KEYWORDS:** Smart Pedestrian Crosswalk, Optimal Location, Big Data, Machine Learning, Geographic Information System(GIS)

**키워드:** 스마트 횡단보도, 최적입지, 빅데이터, 머신러닝, 지리정보시스템(GIS)

### 1. 서론

#### 1.1 연구의 배경 및 목적

최근 어린이 교통사고 문제가 심각한 사회적 이슈로 떠오르고 있다. 2019년 '충남 아산시 어린이 교통사고 사망사건'을 계기로 어린이 교통사고에 대한 사회적·행정적 관심이 증가하여 '민식 이법'은 이례적으로 빠르게 법안 통과되었다. 정부는 도로교통법

개정을 통해 처벌의 수위를 높임으로써 어린이 교통사고를 방지 하려는 노력을 보이고 있으나, 도로교통법이 차량 위주로 설계됐다는 점에서 근본적인 문제 해결이 힘들다(Park, 2021).

어린이 교통사고는 실제 시민들의 관심이 집중된 사안이라는 점에서도 해결이 시급하다고 할 수 있다. 데이터분석 전문업체 RSN 제공 소셜미디어 정보량 분석에 따르면, 검색 키워드 "교통 사고"의 연관 키워드 중 "인물/사람" 관련 키워드의 약 26.3%가

<sup>1)</sup> 학생회원, 한국외국어대학교 경영학과 (sh9213@hufs.ac.kr)

<sup>2)</sup> 학생회원, 한국외국어대학교 경영학과 (sssywon@hufs.ac.kr)

<sup>3)</sup> 학생회원, 한국외국어대학교 경영학과 (ksi0634@hufs.ac.kr)

<sup>4)</sup> 정회원, 홍익대학교 도시공학과 교수 (jkleel@hongik.ac.kr)

<sup>5)</sup> 정회원, 한국외국어대학교 경영학과 교수 (wyun@hufs.ac.kr) (교신저자)

교통약자(어린이, 장애인, 환자)에 해당하였다. 특히 본 논문의 연구대상 지역인 창원시의 경우, 검색 키워드 “창원 교통사고” 관련 키워드 중 “인물/사람” 관련 키워드에서 “어린이” 관련 키워드가 차지하는 비중은 약 60.5%에 달했다.

이처럼 심각한 사회적·행정적 사안인 어린이 교통사고 문제를 해결하기 위해서는 법적인 보완 외에도 보행환경 개선이 우선되어야 한다. 교통사고분석시스템(TAAS)에 따르면 어린이 교통사고 중에서 ‘보행 중 사망’이 66.7%에 달하는 실정이다. 이는 어린이 보행환경의 개선이 전체 시도에서 급급히 다루어야 하는 사안임을 시사한다. 특히, 창원시의 경우 경상남도 평균 대비 ‘차대 사람’ 사망 비율이 가장 높으며, 특히 ‘횡단 중’ 비중이 가장 높았다 (TAAS 2021).

정부는 이러한 교통사고의 심각성을 인지하고 지난해 7월 ‘한국판 뉴딜 정책’에서 ‘국민안전 SOC 디지털화’를 10대 추진과제로 선정했다. 또한 ‘안전하고 효율적인 교통망 구축’을 위해 인공지능(AI) 및 디지털 기반의 관리체계를 도입하여 도로 이용자의 ‘안전과 편의’를 높이는 것을 목표로 제시했다. 뿐만 아니라 지자체들은 ‘스마트 횡단보도’와 같은 첨단 스마트 기술 도입을 통해 어린이 보행 중 교통사고 문제를 해결하고자 노력하고 있다. 스마트 횡단보도란 보행자의 특성을 고려한 시청각 시설을 통해 보행안전 경각심을 높이고 사고를 방지하는 시설물을 말한다 (Seoul, 2016). 실제로 2018년 스마트 횡단보도를 도입한 전라남도의 경우, 교통사고 사망자 수가 도입 전후로 27.4%(106명) 감소한 바 있다. 이에 다른 지자체에서도 스마트 횡단보도의 설치가 적극적으로 논의되고 있다. 하지만 정책 집행에 필요한 객관적인 데이터와 인력이 부족하다는 한계점이 존재한다. 이러한 한계점을 극복하는 동시에 정책적 논의를 할 수 있는 연구는 바로 최적입지 선정에 대한 것으로, 이를 통해 어린이 교통사고 예방 효과를 높이는 것은 물론 효율적인 예산 집행을 기대할 수 있다.

국도교통부는 공간 빅데이터 분석 플랫폼 표준분석모델 중 하나로 스마트 횡단보도 입지를 선정한 바 있다(Ministry of Land, Infrastructure and Transport, 2020). 이는 분석 대상 지역의 횡단보도를 기준으로 유동인구, 학교와의 거리, 사고건수 등을 각각 등급화하여 점수로 환산하고 가중치를 부여한 값을 종합한다. 이를 통해 횡단보도별 점수를 산출하여 최적 스마트 횡단보도 후보 지역을 도출하는 방식이다(Ministry of Land, Infrastructure and Transport, 2020). 하지만 위의 예는 스마트 횡단보도 수요를 산정하는데 필요한 교통사고 혹은 어린이 관련 시설 등의 변수를 고려하지 못했다는 한계점이 존재한다. 서울시 성동구청 또한 스마트 횡단보도 최적입지 선정을 위해 스마트 횡단보도 입지 지수를 산출한 바 있다(Seongdong-gu, 2020). 지수 산출을 위해 횡단보도 수요지수, 교통사고 위험지수, 어린이 교통포용지수 및 고령자 교통포용지수를 모델에 포함시켰다. 성동구청의 경우

국도교통부에 비해 보다 다양한 변수를 포함했다는 점에서 그 의의가 있다. 하지만 여러 변수 간의 가중치를 고려하지 못했다는 점과 다양한 입지 중 각 입지별로 적합한 시설물이 무엇인지에 대한 논의로 발전시키지 못했다는 한계점이 존재한다.

## 1.2 연구의 범위 및 방법

본 연구는 어린이를 고려한 스마트 횡단보도의 최적입지 선정 방안을 제시하고자 한다. 이를 위해 다양한 공공 데이터를 활용하고 머신러닝을 기반으로 하는 분석을 통해 어린이 스마트 횡단보도 최적입지를 선정하고자 한다. 공적 영역의 입지 문제는 사회적 관점에서 공공의 목적을 달성하고자 한다는 점에서 상업적 이익을 최대화하려는 사적 영역과는 다른 특징을 지니고 있기 때문에 본 연구의 의의가 있다. 또한, 어린이 교통사고가 심각한 사회적·행정적 문제로 여겨지는 창원시를 대상으로 하여 추후 보행환경 개선에 기여하고자 한다.

앞서 설명한 선행 분석의 한계점을 고려, 본 연구는 어린이 이용 시설과의 접근성, 인구수 등의 변수를 포함하여 어린이 교통사고 예측모형을 개발하였다. 해당 모형을 통해 어린이 교통사고에 영향을 미치는 변수들을 선정하고 그에 따른 가중치를 부여하여 최종 입지 지수를 산출하였다. 또한, 본 연구는 단순한 최적입지 선정에서 그치는 것이 아니라 k-means clustering을 통해 최적입지를 유사한 것끼리 유형화함으로써 각 입지에 적합한 시설물을 제안하고자 한다. 그 결과, 각각 낮 시간대 보행자 중심, 저녁 및 밤 시간대 보행자 중심, 운전자 중심의 스마트 횡단보도 최적입지로 유형화되었고 각 유형에 적합한 시설물을 제안한다. 이러한 분석을 통해 불필요한 스마트 횡단보도 시설물의 설치를 방지함으로써 행정 예산의 효율적 운용에 기여하고 스마트 횡단보도의 효과를 극대화하고자 했다. 이러한 본 연구는 스마트 횡단보도 및 교통안전과 관련된 연구의 폭을 넓히는데 기여했으며 정책 설계에 있어서 공공 데이터 및 개방 데이터에 근거한 의사결정을 도입한다는 의의가 있다. 이러한 연구의 시사점을 고려해보았을 때, 본 연구에서 중심으로 다루는 어린이 외에도 다양한 취약계층의 교통안전에 향상하는데 기여할 수 있을 것으로 기대된다.

본 논문의 구성은 다음과 같다. 2장에서는 공공시설 입지선정 관련 이론적 배경을 고찰했으며, 3장에서는 활용 데이터 및 분석 방법을 제시하였다. 4장에서는 머신러닝을 활용한 입지 지수 및 이를 기반으로 창원시 스마트 횡단보도의 최적입지 선정 결과를 제시하였다. 5장에서는 k-means 클러스터링을 활용한 입지의 유형화 및 유형별 시설물을 제안했고, 6장에서는 본 연구에 대한 결론과 향후 연구 과제를 도출하였다.

## 2. 이론적 배경

### 2.1 최적입지선정

공간지리(GIS)분석 분야에서 공공시설물의 최적입지 선정은 오랫동안 중요하게 다뤄진 연구 주제이다. 상업적 이익의 극대화를 고려하는 사적 건물과는 달리, 공공시설물은 효율성 그리고 형평성과 같이 서로 상충하는 개념들을 입지선정 과정에 포함해야 한다는 점에서 다기준 평가방법이 선호되어온 바 있다(Lee et al., 2019). 하지만 다기준 평가방법은 비공간적 평가 기준에 관한 접근법이라는 점에서 공간적으로 균등하지 않게 분포된 입지 기준에 대해 적절하지 못하다는 한계점이 있다(Lee, 2000). 이러한 공간적 한계를 극복하기 위해 최근에는 GIS를 활용한 연구들이 진행되고 있다. 공공시설 최적입지에 관한 연구의 경우, 초기에는 주로 행정시설물 혹은 남비시설의 입지선정에 관한 연구가 주를 이뤘다(Lee, 2000; Kim and Chung, 2001; Kim and Kim, 2006). 하지만 최근의 연구들은 자전거 주차장, 사전투표소, 무인택배함, 스마트 버스정류장 등과 같이 시대적 흐름에 따라 새로이 생겨난 시설물에 관한 입지선정 모델링 연구와, 재난·안전에 관한 주제가 주를 이루고 있다(Choi and Park, 2018; Park et al., 2017; Song and Lee, 2017; Lee et al., 2017; Choi et al., 2021). 또한, 공간 빅데이터의 수집 및 분석이 용이해지면서 최근 머신러닝 기법을 활용한 보다 정확한 입지선정 방법론에 관한 관심도 높아지고 있다(Lee and Lee, 2016; Yoon and Shim, 2021).

### 2.2 어린이 교통사고

#### 2.2.1 어린이 교통사고의 심각성

어린이를 비롯한 교통 취약계층에 대한 시민들의 관심은 상당하다. 특히 어린이 교통사고는 정치, 경제, 사회 등 다양한 분야에서 민감하게 다뤄지고 있는 주제이다. 그중 정치·사회 분야에서 '민식이법'을 계기로 어린이 교통사고의 심각성이 제고된 바 있다. 이는 2019년 9월 충청남도 아산시의 어린이보호구역에서 교통사고를 당한 김민식군 사건 직후 발의된 법안이다. 해당 사고는 학부모를 비롯한 다수의 시민들에게 공분을 샀다. 한국도로교통연구원(2021)에 따르면, 도로교통사고로 인한 사회적 비용은 43.3조원에 달하며, 이는 2018년보다 1.6조원이 증가한 규모이다.

#### 2.2.2 어린이 교통사고 관련 선행연구 고찰

어린이 교통사고의 심각성 및 사회적 인식을 반영하여, 학계에서는 교통사고에 관한 연구가 진행된 바 있다. Kang et al., (2016)은 서울에 거주하는 어린이 보행자 도로교통사고와 위험요인을 분석했는데, 단계적 회귀분석(Stepwise Regression)을 활용하여 어린이 보행자 교통사고 특성을 검토한 바 있다. Yoon et al., (2016)는 순서형 로지스틱 회귀분석(Ordered Logistic

Regression; OLR)을 활용하여 어린이 교통사고 심각도에 영향을 미치는 주요 요인들을 발견하고 이를 제시했다. Lee and Woo, (2018)는 통계적 방법을 이용한 추정모형과 일반화 회귀 신경망(Generalized Regression Neural Network; GRNN)과 같은 머신러닝을 이용한 추정모형의 적용성, 설명력, 유연성을 비교했다. Lee and Lee, (2021)는 보행자 교통사고 사고 심각도 예측에 다양한 유형 분류 머신러닝 알고리즘을 적용하고 그 성능을 비교하여 교통사고 심각도 예측에 가장 적합한 알고리즘을 도출했다. 서포트 벡터 머신(support vector machine; SVM), 랜덤 포레스트(random forest), 다층 퍼셉트론 (multilayer perceptron), 랜덤 오버 샘플링(random over sampling) 등의 주요 머신러닝 기법을 사용한 바 있다.

어린이 교통사고에 관한 국내 연구는 어린이 보호구역의 실태 연구 및 어린이 교통사고에 영향을 미치는 요인에 관한 연구가 주를 이루고 있다. Park et al., (2017)은 스쿨존 교통안전 개선사업의 시행이 해당 지역의 어린이 교통사고 건수에 미치는 영향을 분석했다. 스쿨존 내 도로의 교통안전 특성과 어린이 교통사고 건수 간의 상관관계를 분석하고자 로지스틱 회귀모형을 이용하여 어린이 교통사고 발생 확률 모델을 개발한 바 있다. Baek et al., (2016)은 어린이 보호구역의 공간적 특성을 고려한 어린이 교통사고모형을 개발하였고 가산자료 모형 중 포아송(Poisson) 모형을 이용했다. 특히, 어린이가 많은 도로에서는 운전자의 주의 및 인지를 향상시킬 수 있는 시설이 필요함을 시사했다. Baek et al., (2013)은 어린이 보호구역 내 어린이 교통사고의 영향 요인을 도출하였으며, 음이항(Negative Binomial) 모형과 RMSE를 이용하여 해당 모형이 통계적으로 유의함을 입증하였다. 모형에 포함된 독립변수는 학생 수, 차로 수, 감시 카메라 수 등이 있다. Na et al., (2021)은 어린이 보호구역 내 회전교차로의 교통사고를 운전 유형별로 분류하여 사고 특성을 분석하고, 사고 건수와 EPDO(Equivalent Property Damage Only)와의 상관관계를 고려하여 보호구역 내의 안전시설에 대한 만족도를 측정하는 연립방정식 모형을 개발하였다. 또한, 운전 유형에 영향을 미칠 수 있는 보호구역 표지 기능에 대한 소비자의 기대치를 설명하기 위해 카노(KANO)모형을 활용하였다.

이처럼 어린이 교통사고 관련 연구는 도로 시설물 관련 변수들에 대한 분석으로 국한된 경우가 대다수이다. 따라서 실제 어린이들의 행동특성 및 생활환경을 고려하지 못했다는 한계점이 존재한다. 또한, 최근 빅데이터와 머신러닝을 활용한 예측모형들을 비교하는 교통사고 관련 연구가 활발히 진행되고 있음에도 불구하고, 어린이 교통사고에 대해서는 머신러닝을 활용한 연구가 부족한 실정이다. 이에 본 연구에서는 머신러닝을 활용한 어린이 교통사고 예측모형을 제시함으로써 선행 연구의 한계점을 보완하는 연구 결과를 도출하고자 한다.

### 2.3 스마트 횡단보도

‘스마트 횡단보도’란 보행자 및 운전자의 특성을 고려하여 다양한 시청각 시설물을 통해 보행안전 경각심을 높이고 사고를 방지하는 시설물을 의미한다 (Seoul, 2016). 보행자의 움직임에 따라 조명이 이동함으로써 운전자가 보행자를 더 잘 인지하게 하는 집중조명, 횡단보도에 대한 보행자의 인지를 효과적으로 환기하는 바닥신호등, 로고라이트 및 음성안내기, 횡단보도에 대한 운전자의 인지 정도를 높이는 정지선 라이트가 그 예이다(Figure 1 참고). 이처럼 스마트 횡단보도는 보행자가 더욱 안전한 보행환경을 누릴 수 있도록 돕는 시설이라고 할 수 있다.

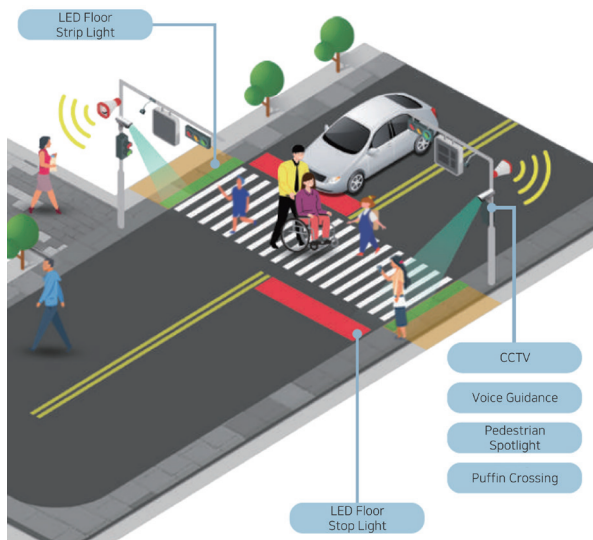


Figure 1. Example of smart pedestrian crosswalk  
출처: 핀텔 (<http://www.pintel.co.kr/>)

앞서 언급하였듯이, 심각한 사회적 문제로 떠오른 도로교통사고 문제를 해결하는 방안으로 스마트 횡단보도와 같은 첨단 ICT 기술 및 스마트 기기의 도입이 활성화되고 있다. 서울시는 2019년 ICT 기술로 교통 및 안전문제를 해결하는 ‘스마트시티 특구’로 성동구를 선정한 바 있다. Lee, (2020)는 성동구의 스마트 횡단보도 설치 1년 후, 차량정지선 위반 건수가 70% 가까이 감소하였음을 밝혔다. 전라남도는 바닥 신호등과 음성 안내기 등의 스마트 횡단보도 91개소를 설치한 바 있는데, 도입 직전 해인 2017년에는 387명이 사망했으나, 2018년에는 281명이 사망하여 약 27.3% 감소치를 보였다(Oh, 2021). 이처럼 스마트 횡단보도는 어린이를 포함한 보행자 교통사고를 해결하는 데 효과적이거나, 대부분의 선행 연구는 시설물 개발과 같은 다소 한정적인 주제에 머물러 있는 것이 현실이다(Jang et al., 2020; An et al., 2016; Jang et al., 2019). 이에 본 연구는 스마트 횡단보도의 최적입지 선정에 대해 논의함으로써, 각 지자체 혹은 정부기관의 효율적인 스마트 횡단보도 입지선정 및 도입에 기여하고자 한다.

## 3. 활용데이터 및 분석방법

### 3.1 연구 범위

본 연구는 경상남도 창원시를 대상으로 어린이를 고려한 스마트 횡단보도 최적입지를 선정하였다. 창원시의 경우, 타 시도와 비교하면 보행자 교통사고 문제가 심각한 실정이다. 교통사고분석시스템(TAAS, 2018)에 따르면, 창원시의 보행자와 교통약자에 대한 교통안전지수는 최하등급(E)으로 전국 227개 지자체 중 최하위의 점수를 받아 타 시도에 비해 심각하게 낮은 상황이다. 이에 창원시는 교통사고 사망자와 사고 건수를 줄이기 위해 교통안전시설 개선, 제한속도 하향 사업, 그리고 스마트 횡단보도와 같은 최신 기술과 교통사고 개선 방안을 새롭게 도입하였다. 하지만 2019년에도 창원시는 전국 지자체 중 교통안전지수 총점 및 개선율이 최하위 20%에 포함되면서 실효성 있는 교통정책의 대대적인 도입이 필요한 상황이다 (TAAS, 2019).

창원시 교통사고 중 가장 시급하게 개선이 필요한 분야는 “어린이 보행 중 교통사고”이다. 사고유형별 교통사고(TAAS, 2020) 데이터에 따르면, 창원시 교통사고 중 차대사람 사고 1위 유형은 “횡단 중 사고”인 것으로 나타났다. 특히 어린이 교통사고 34,415건 중 37.6%가 보행(횡단)중 사고인 것으로 나타났다(TAAS, 2018). 현재 창원시는 스마트 횡단보도 도입 초기단계에 있다. 일부 어린이 보호구역과 초등학교를 중심으로 시범운영 혹은 도입을 논의 중이다. 이에 본 연구는 창원시의 스마트 횡단보도 정책의 적절성 평가와 추후 이뤄질 스마트 횡단보도 확대 운영에 큰 기여를 할 수 있을 것이다.

### 3.2 데이터 수집

본 연구를 수행하기 위해서 공간지리데이터 등 다양한 2차 자료를 수집하였다. 분석에 활용한 데이터는 모두 공공 데이터를 기반으로 수집·가공된 것이다. 교통사고분석시스템(TAAS)에서 2012년부터 2020년까지의 교통사고 위치정보와 사고 정보를, 창원시 빅데이터 포털에서 창원시 신호등, 초등학교, 유치원 및 어린이집, 버스정류장 위치정보 등을 각각 수집하였다. 또한, 국가통계포털, 공공 데이터 포털에서 각각 행정동별 인구 정보 및 학원과 교습소 위치정보를 수집하였다. Table 1에서 활용한 데이터 목록과 형태, 출처에 대한 자세한 설명을 기술하였다.

Table 1. Data type and source

Data	Type	Source
Traffic light location	Point	Changwon big data portal
Traffic accident information	Polygon/Point	The road traffic authority
Population	Polygon	Korean statistical Information service
Elementary school location	Point	Changwon big data portal
Kindergarten location	Point	Changwon big data portal
Private institute location	Point	Public data portal
Bus stop location	Point	Changwon big data portal

### 3.3. 전체 분석 프로세스 개괄

본 연구는 어린이를 고려한 스마트 횡단보도의 최적입지 선정 방안을 제시하고자 한다. 앞서 설명한 선행 분석의 한계점을 고려, 다양한 머신러닝 기법과 공공 데이터를 활용하여 어린이 생활 특성을 고려한 교통사고 예측모형을 개발하였다. 해당 모형을 통해 어린이 교통사고에 영향을 미치는 여러 변수 중 의미 있는 변수들을 선정하고 그에 따른 가중치를 부여하여 최종 입지지도 산출하였다. 해당 입지지도를 활용하여 1,036개 그리드로 구분한 창원시 격자(300m×300m)에 적용한 뒤, 최적입지를 선정했다. 단순히 입지를 선정하는 데 그치지 않고, 본 연구에서는 선정된 횡단보도 입지들의 특성 및 클러스터링을 통해 유형화할 것이다. 전체 연구 프로세스는 Figure 2와 같다.

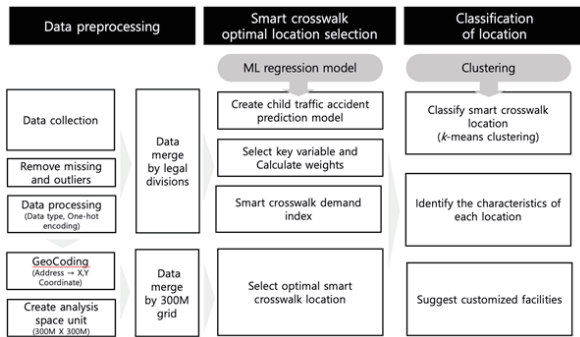


Figure 2. Scope and method of research

## 4. 머신러닝 변수 중요도를 활용한 입지 지수 제안

본 연구는 기존의 입지선정 방법의 한계점을 파악하고, 지수 산출의 과학적 근거를 마련하기 위해 머신러닝을 이용하여 스마트 횡단보도 입지선정 기준을 정의하고자 한다. 이를 위해 본 연구에서는 어린이 교통사고 예측에 있어서 가장 높은 정확도를 보인 모델을 선정한 뒤, 해당 모델이 제공하는 변수중요도를 통해 주요 변수와 가중치를 산출하는 방식으로 최적입지 지수식을 산출했다.

### 4.1. 머신러닝 모델 개발 및 비교

본 연구에서는 어린이 교통사고 예측모형을 개발하기 위해 총 50개의 설명변수가 제안되었는데, 선행 연구에서 사용된 변수들을 기반으로, 어린이들이 많은 시간을 보내거나 자주 방문하는 장소 그리고 시설물 및 교통사고와 관련된 변수들을 추가했다(Ministry of Land, Infrastructure and Transport, 2020; Seongdong-gu, 2020). 본 논문에서 정의하는 어린이는 도로교통법에 따라 만 13세 미만으로 규정했다. 교통사고 발생 정보(날

짜, 시간, 날씨 등), 인구밀도 및 이동량, 어린이 관련 시설물(초등학교, 어린이집, 학원 및 교습소, 공원 등) 등을 독립변수로, 행정동별 어린이 교통사고 사고 건수를 종속변수로 설정하고 모델을 개발하였다. 본 연구에서는 90개의 창원시 행정구역을 기준으로 데이터를 분석하였으며, 일반구를 기준으로 집계한 기초 통계량은 Table 2와 같다. 또한 독립변수들의 기초 통계량은 Table 3와 같다.

Table 2. Descriptive statistics for child traffic accidents

District	Masan happo	Masan hoewon	Seongsan	Uichang	Jinhae
Count	32	8	6.4	15	15
Mean	5.3	17.8	6.4	10.7	13.9
Std	6.6	13.6	6.4	9.5	10.6
Min	0	1	1	1	1
25%	1	11.4	1	2.5	5.5
50%	2	14	3.5	6	13
75%	6.5	22.8	10.3	16.5	19.5
Max	23	39	21	20	33

Table 3. Descriptive statistics

Variable	Count	Mean	Std	Min	25%	50%	75%	Max
Number of Accidents	90	9.0	9.4	0	1	5	14	39
Teenage Population	90	7.3	1.9	0	7	7.7	8.4	9
Elementary School	90	2.3	1.5	0	1	2	3	8
Private Institute	90	50.9	45.4	0	18	30	80	202
Kindergarten	90	3.7	2.8	0	2	3	5	13
Total Population	90	9.9	0.7	8.2	9.4	9.9	10.4	11.1
Bus Stop	90	46.0	33.3	4.0	26.3	35	56	205

본 연구에서는 입지 지수 계산에 적용할 가중치 산출을 위해, 종속변수를 행정동별 어린이 교통사고 사고 건수로 두고 여러 머신러닝 모형을 적용하였다. 모델의 평가를 위해 손실 함수로 사용되는 평균 제곱근 오차(Root Mean Square Error; RMSE)를 이용함으로써 예측 성능을 평가하고자 했다.

전통적인 통계모형인 Linear Regression을 기본으로, 해당 모델에서 파생된 Lasso Regression, Ridge Regression, Elastic Net 모델과 최신 머신러닝 기법의 하나인 LGBMRegressor를 사용하였다. Lasso Regression은 많은 변수가 있을 때, 영향력이 큰 변수만 남기고, 영향력이 적은 변수들의 계수를 0으로 만든다. 변수의 수가 줄어들어 과적합(Overfitting)을 방지하는 장점이 있다. 하지만 변수 간의 상관관계가 있을 때 그중 하나의 변수만 채택하고 나머지 변수의 계수를 0으로 만들기 때문에 정확도가 낮아지고, 정보가 손실될 수 있다. Ridge Regression도 Lasso Regression 유사하게 계수들을 0에 가깝게 만든다. 이로 인해 과적합을 방지하는 기능은 조금 떨어지지만, 정확도를 높이고 손실을 줄일 수 있다. Elastic Net은 Lasso와 Ridge 모두를 사용하기 때문에 두 Regression의 장점을 모두 가지고 있다. 영향력이

큰 변수는 채택하고 영향력이 작은 변수는 0에 가깝게 만든다. 각 모델의 RMSE를 비교해본 결과, 전통적인 통계 모델보다 머신러닝 모델인 LGBMRegressor의 RMSE가 가장 낮게 나타났다 (Figure 3).

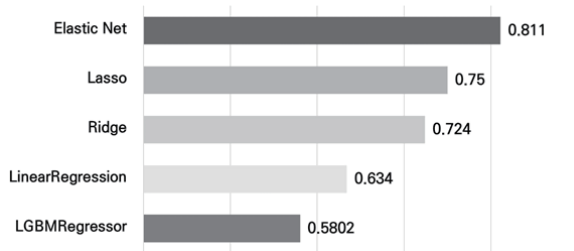


Figure 3. RMSE by machine learning prediction models

## 4.2. 최종 선정 모델

최종적으로 선정된 LGBMRegressor의 경우, 트리 기반의 LGBM 모델을 회귀모형에 적용한 것이다. LGBM 앙상블 학습 (Ensemble Learning)이란 여러 개의 분류기를 생성하고, 그 예측을 결합하여 더 정확성이 높은 예측을 도출하고자 하는 기법이다 (Ke et al., 2017). 다양한 앙상블 학습 유형 중에서 boosting은 여러 개의 분류기가 순차적으로 학습을 진행하며, 예측이 맞지 않은 데이터에 대해서 보다 정확하게 예측하는 것이 가능하도록 다음 분류기에 가중치를 더하면서 진행되는 방식이다.

기존의 앙상블 모델과 LGBM은 다음과 같은 차이가 있다. 기존의 앙상블 모델은 트리의 깊이를 감소시키기 위해 level-wise 하도록 수평적으로 증가하며 분할되는 반면, LGBM은 트리의 균형을 맞추지 않는 형태로 leaf-wise 하게 수직적으로 증가하며 분할된다. 이 경우, 동일한 leaf를 생성했을 때, level-wise에 비해 leaf-wise는 손실이 줄어든다는 특징이 있다 (LightGBM 2020; Geron, 2018; Friedman, 2001).

이처럼 LGBM 모델의 장점 때문에 다양한 분야에서 분류 및 예측모형으로써 활용되고 있다 (Kim et al., 2019; Oh et al., 2021; Kim et al., 2020). 특히 LGBM은 GOSS (Gradient Based One Side Sampling)와 EFB (Exclusive Feature Bundling) 방식을 활용하여, 기존 Gradient boosting 기법에서 고차원 변수의 데이터가 큰 경우 발생하는 낮은 효율과 확장성을 개선하였다. 즉, 적은 양의 데이터로도 정확하게 정보 획득을 추정할 수 있다는 장점이 있다. 본 연구의 데이터는 다차원의 독립변수로 구성되며, 데이터 표본이 독립변수의 개수에 비해 상대적으로 적다는 특성을 가지고 있다. LGBM 모델의 작동방식과 본 연구에서 사용되는 데이터의 특성을 고려한다면, LGBM은 연구모델로 적합하다고 할 수 있다.

## 4.3. 최적입지 지수식 산출

본 연구에서는, 스마트 횡단보도 수요의 정도를 하나의 숫자로 표현하여 지역 간 비교가 용이하도록 입지 지수 지표를 계산하여 사용한다. 해당 지수를 개발하기 위해 Gan et al., (2019)의 방법론을 차용하여 머신러닝 변수중요도를 활용한 지수를 개발하였다.

Figure 4는 LGBM 모델 실행결과 산출된 변수중요도 지수 상위 20개 변수를 시각화한 그림이다. 회귀 계수 시각화 결과, '버스 정류장 수'가 어린이 교통사고에 많은 영향을 끼쳤고, 다음으로 '학원 수'가 어린이 교통사고에 많은 영향을 끼친 것을 확인할 수 있다. 특히, 학원 및 교습소의 경우는 어린이들이 많이 상주하는 시설물이지만, 어린이 보호구역에는 포함되지 않기 때문에 어린이 교통사고의 사각지대에 놓여 있다. 이는 학원 및 교습소 또한 입지 선정 기준에 포함되어야 함을 시사한다.

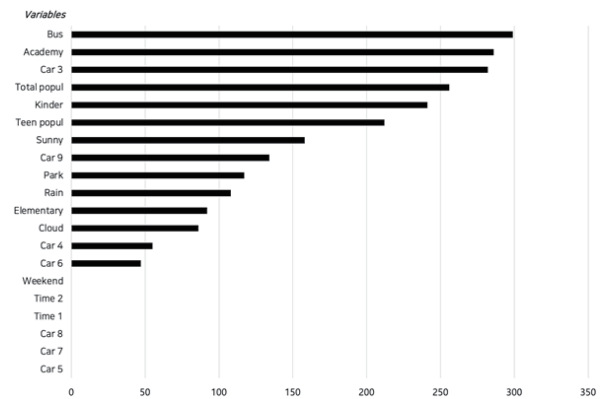


Figure 4. LGBMRegressor feature importances

추출된 상위 20개 변수 중에서 날씨나 가해자의 차종과 같이 유동적이면서, 안전 시설물의 설치로 사고 예방이 어려운 변수들은 (13개) 분석에서 제외된 뒤, 나머지 6개 변수의 변수중요도를 가지고, 변수 가중치를 산출했다. 선정된 6개 변수에 대한 분포도는 Figure 5와 같다. 부여할 가중치는, 가장 높은 중요도를 가지는 변수의 중요도로 해당 변수의 중요도를 나눈 값에 절대값을 취한 다음, 어린이 교통사고 건수와와의 상관관계에 따라 양수, 혹은 음수의 부호를 취하여 사용하였다. 예를 들어, 가장 높은 변수 중요도 지수를 가지고 있는 버스 정류장 수의 변수중요도 지수가 299이고, 여러 변수 중 하나인 총인구수의 변수중요도가 256일 때, 총인구수 변수에 대해 부여할 가중치의 절대값은  $256/299 = 0.86$ 이며, 총인구수와 어린이 교통사고는 양의 상관관계를 가지고 있으므로 양(+)의 부호를 붙여주는 방식으로 가중치를 설정하였다. 이러한 방식으로 계산되는 입지 지수를 창원시 그리드에 적용하여 각 그리드에서의 개별 입지 점수를 산출하였다 (Gan et al., 2019).

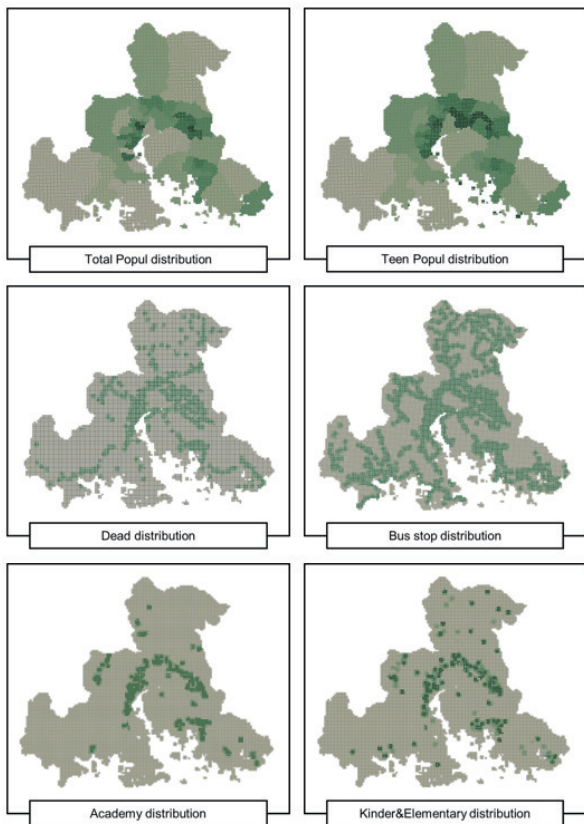


Figure 5. Variable distribution

#### 4.4. 최적입지 선정

창원시 내에서도 스마트 횡단보도가 우선적으로 도입되어야 할 지역을 찾아내기 위해서 QGIS 프로그램을 활용하여 창원시를 1,036개 신호등 그리드로 구분하였다. 어린이 보호구역 선정 기준이 반경 300m 이내라는 사실과 창원시 내 횡단보도가 격자 내에 1~2개 정도 고루 분포할 수 있다는 점을 고려하여, 창원시를 300m×300m 격자로 나누어 그리드별 입지점수를 산출했다. 창원시 횡단보도 격자 중 어린이 교통사고 위험이 높아 보행자 보호장치 도입이 시급하다고 판단되는 입지점수 상위 10%에 해당

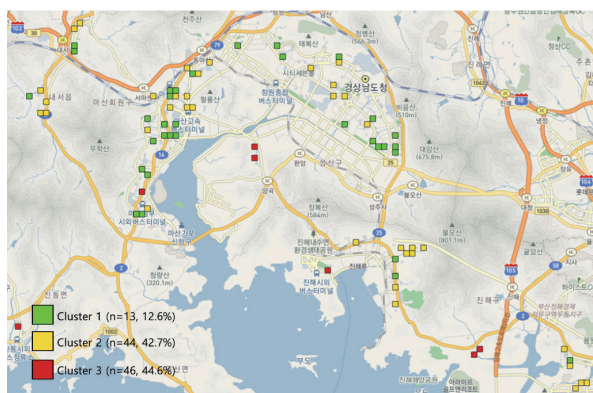


Figure 6. Optimal location

하는 103개의 격자를 입지 우선 지역으로 선정했다(Figure 6). 본 분석이 제시한 스마트 횡단보도 최적입지는 창원시가 실제 시범 운영 중인 스마트 횡단보도 위치와 다수 겹치는 것을 확인할 수 있었다. 그 예로 도계초등학교 인근 및 유니시티 아파트 인근 등이 있다. 이는 본 분석이 정확하고 체계적으로 이루어졌으며, 추후 창원시 교통정책에 적용될 가능성이 높음을 시사한다.

### 5. k-means 클러스터링 모델을 활용한 최적 입지의 유형화 결과 및 유형별 시설물 제안

본 연구에서는 예산의 효율적 활용과 정책 효과의 극대화를 위해 최적입지 선정 후, 입지별 특성을 파악하여 이를 유형화한 뒤 각 입지에 적합한 시설물을 제안하고자 한다. 본 연구에서는 입지 유형화를 비지도 학습(Unsupervised learning)기법 중의 하나인 k-means 클러스터링을 활용하였다.

#### 5.1 최적 클러스터 수 도출

다수의 선행 연구에서 입지 유형화를 위해 활용되고 있는 k-means 클러스터링 분석은, 클러스터 내 개체 유사성과 클러스터 간 개체 차이성이 최대화되는 독립적 클러스터 분석의 일종으로, 종속변수가 배타적인 방식으로 군집화되어 특정 변수가 하나의 명확한 클러스터에 속하도록 군집이 분류되는 특성이 있다(Yun et al., 2017; Yeom and Kim, 2011; Pejman et al., 2017). 클러스터링을 이용하여 입지를 유형화하기 이전, 값의 범위가 큰 변수가 변수중요도와는 상관없이 더 큰 영향을 미치게 되는 경우를 방지하기 위해, 연속형 변수들의 값을 Min-Max Standardization을 이용하여 0과 1 사이의 숫자 값으로 스케일링 하였다. 이후 위에서 언급한 데이터를 가공하여 그리드별 통계량을 산출하였고, 이를 바탕으로 Python을 활용하여 클러스터링 분석을 진행하였다. 이후 각 클러스터의 특성에 대한 분석을 진행하였다.

본 연구에서는 최적의 클러스터 수를 도출하기 위해 유클리드 거리 제곱 값을 사용하여 2개부터 10개의 클러스터를 생성하였다. 이후, Elbow, silhouette 계수를 비교하여 최적의 클러스터 수를 판단한 결과, k가 3과 5일 경우가 가장 최적의 클러스터 수로 분석되었다(Figure 7). 실루엣 계수(Coefficient)는 클러스터가 얼마나 잘 분류되었는지를 평가할 수 있는 지표이다(Rousseeuw, 1987). 실루엣 계수는 -1부터 1까지의 값을 갖는데, 값이 1에 가까울수록 다른 클러스터들과는 거리가 멀고, 동일한 클러스터 내의 데이터끼리는 서로 가깝게 잘 뭉쳐 있다는 것을 의미한다(Lim and Lee, 2021).

한편, k값이 작으면 일반적으로 클러스터 생성에 유리하고,

103개라는 적은 표본의 수를 고려할 때, 본 연구에서는  $k=3$ 으로 설정하여 클러스터 분석을 진행하였다. 또한, 스마트 횡단보도의 초기 도입단계에서는 개별 지역적 특성의 반영보다는 거시적인 환경 파악이 중요하다고 판단하여, 비교적 포괄적인 지역 특성이 반영된 3개의 클러스터 분류를 사용하는 것이 적합하다고 판단하였다.

선정된 횡단보도 입지들의 특성 및 클러스터링을 통한 유형화 결과, (1) 낮 시간대 보행자 중심의 스마트 횡단보도 최적입지, (2) 운전자 중심의 스마트 횡단보도 최적입지, (3) 밤 시간대 보행자 중심의 스마트 횡단보도 최적입지 등으로 유형화되었다. 각각의 클러스터에 대한 설명과 적절한 시설물에 제안은 아래에 더 자세히 설명한다.

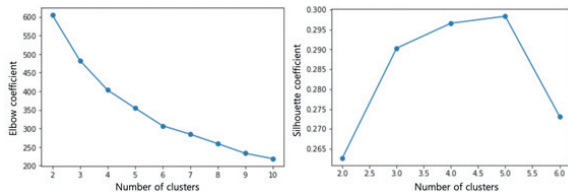


Figure 7. Elbow and silhouette coefficient

## 5.2 클러스터 별 특성 파악 및 시설물 제안

### 5.2.1. 시설물 설명

본 연구에서 제안하는 클러스터별 시설물은 관제 CCTV, 정지선 라이트, 집중조명, 음성안내기, 바닥 신호등, 로고라이트, 보행 신호 자동연장시스템(Puffin crossing)이다. 관제 CCTV는 횡단보도 주변을 녹화하는 시설물로서, 운전자에게 횡단보도에서의 안전 운전을 유의시키는 효과가 있다. 정지선 라이트는 기존에 페인트로 표시되었던 횡단보도에 인접한 차도 정지선에 LED 조명 시설물을 설치한 것이다. 정지선 라이트는 교통신호에 따라 점멸하며, 이를 통해 운전자는 정지선에 대한 인지 정도가 높아져 정지선을 준수할 수 있다. 집중조명은 보행자의 이동에 따라 움직이며 보행자를 비추주는 조명을 의미한다. 이는 특히 야간에 횡단보도를 이용하는 보행자를 운전자에게 더 쉽게 인지시킬 수 있다는 이점이 있다. 음성안내기는 음성으로 신호 정보를 안내하는 시설물을 의미하며, 바닥 신호등은 횡단보도 시작 지점의 인도 바닥에 설치된 LED 신호등을 의미한다. 이를 통해 보행자는 더욱 분명하게 교통신호를 인지할 수 있다. 로고라이트는 횡단보도 주변 인도에 설치된 LED 시설물로, 보행자의 횡단보도 존재를 인지시키는 역할을 한다. 마지막으로, 보행 신호 자동연장시스템은 카메라를 활용하여 보행자의 필요에 따라 횡단보도 보행을 선택할 수 있으며, 교통약자가 직접 횡단시간을 연장할 수 있도록 지원한다. 이를 통해 교통약자를 비롯한 보행자들은 충분한 시간을 갖고 안전하게 횡단할 수 있는 효과가 있다.

### 5.2.2. 클러스터별 특징 및 스마트 횡단보도 제안

Table 4. Cluster characteristics and suggested features

Avg Score per Cluster	Cluster 1	Cluster 2	Cluster 3
Total popul	0.403	0.893	0.904
Teen popul	0.472	0.879	0.907
Bus stop	0.122	0.129	0.071
Dead	0.395	0.303	0.22
Elementary	0.307	0.344	0.207
Kinder	0.045	0.008	0.065
Academy	0.046	0.008	0.065

Classified Clusters			
Cluster	(1) Heavy traffic	(2) School zone	(3) Academy& housing complex
Description	Driver-centric smart crosswalk	Daytime smart crosswalk	Day&night smart crosswalk
Suggested crosswalk features	CCTV	CCTV	LED floor strip light
	LED floor stop light	LED floor stop light	Pedestrian spotlight
	Pedestrian spotlight	Voice guidance	Logo light
	Puffin crossing	Puffin crossing	Puffin crossing

Table 4는 3개의 클러스터별 특징과 각각의 클러스터에 적합한 스마트 횡단보도 디자인을 제안한다. 클러스터 1의 경우, 어린이 주거인구와 전체 주거인구 비율은 다른 지역에 비해 낮으나 마산역, 마산 고속버스터미널, 마산 시외버스터미널 등의 교통거점 및 여가시설이 집중되어 있었다. 이로 인해 해당 클러스터는 다른 클러스터에 비해 교통량이 많았다. 따라서 해당 클러스터에는 ‘운전자 중심의 스마트 횡단보도’가 설치될 필요가 있다는 분석 결과가 도출되었다. 이러한 클러스터 1의 특징을 고려하여, 해당 클러스터에는 관제 CCTV, 정지선 라이트, 집중조명을 설치할 필요가 있다. 관제 CCTV는 운전자가 횡단보도 근처를 주행할 때 교통사고에 대한 경각심을 가지고 차량을 운전하는 데 도움을 줄 것으로 기대된다. 정지선 라이트는 운전자의 정지선에 대한 인식 정도를 높임으로써 보행자의 안전을 보장하는 교통 법규의 준수를 도울 수 있을 것이다. 마지막으로, 집중조명은 야간에 횡단보도를 이용하는 어린이 보행자가 운전자에게 더 잘 인식되게 함으로써, 운전자는 어린이 보행자의 안전에 더욱 주의하여 차량을 운행할 수 있을 것으로 기대된다.

클러스터 2의 경우, 어린이 주거인구와 총 주거인구 비율이 클러스터 3에 비해 낮으나 약 16개의 학교가 밀집된 지역에 해당했다. 즉, 등하교 시간대 어린이 인구 이동량이 매우 높을 것으로 추정되는 곳이다. 따라서 해당 클러스터에는 등하교 시간대에 해당하는 ‘낮 시간대에 맞춰진 보행자 중심의 스마트 횡단보도’ 도입이 필요하다는 분석 결과가 도출되었다. 이러한 클러스터 2의 특징을 고려하여, 해당 클러스터에는 음성안내기, 정지선 라이트, 관제 CCTV를 설치할 필요가 있다. 음성안내기는 청각적 요소를 통해 어린이 보행자가 횡단보도 및 교통신호를 인지하게 함으로써 안전에 더욱 유의하며 횡단보도를 건너게 할 수 있다. 정지선



라이트와 관제 CCTV 또한 운전자가 어린이 보행자에 유의하며 운전을 하도록 유도한다. 이러한 횡단보도를 조성함으로써 보다 안전한 보행환경을 도모할 수 있을 것으로 기대된다.

클러스터 3의 경우, 어린이 주거인구 및 총 주거인구 비율이 높고, 주거단지 및 학원이 밀집되어 있다. 때문에 클러스터 1, 2에 비해 저녁 및 밤 시간대의 어린이 이동량이 많은 것으로 분석되었다. 대부분 어린이 보호구역으로 지정된 클러스터 2(예: 학교 근처 횡단보도)와 달리 학원이 근처는 어린이 보호구역으로 지정되지 않은 경우가 많다. 이 때문에 어린이를 위한 보행자 시설물이 미비한 실정이다. 이에 어린이들이 학원을 이용하는 '저녁 및 밤 시간대에 맞춰진 보행자 중심의 스마트 횡단보도' 도입이 필요하다는 분석 결과가 도출되었다. 이러한 클러스터 3의 특징을 고려하여, 해당 클러스터 횡단보도에는 바닥 신호등, 집중조명, 그리고 로고라이트를 설치할 필요가 있다. 로고라이트와 바닥 신호등은 보행자가 야간에도 횡단보도 및 신호등에 대한 인식 정도를 높임으로써 어린이 보행자가 안전에 더욱 유의하여 횡단보도를 건너는데 기여할 수 있다. 또한, 집중조명은 어린이 보행자의 움직임에 따라 함께 이동하기 때문에, 어린이 보행자의 이동 경로를 분명히 함으로써 보다 안전한 환경에서 횡단보도를 건너는데 기여할 수 있을 것이다.

마지막으로, 클러스터 1, 2, 3 공통으로 보행 신호 자동연장시스템을 설치할 필요가 있다. 보행 신호 자동연장시스템은 보행자의 횡단을 기준으로 신호 시간을 조절하며, 보행자가 직접 횡단 시간을 연장할 수도 있다. 이를 통해 성인보다 비교적 긴 횡단시간이 필요한 어린이 보행자가 충분한 시간을 가지고 안전하게 횡단보도를 건널 수 있을 것으로 기대된다.

## 6. 향후 연구 진행 방향 및 결론

본 논문에서는 스마트 횡단보도의 최적입지 선정 방법을 제안하기 위해 머신러닝 기법을 이용하여 최적입지 지수산출법을 개발했다. 먼저, 창원시를 1,036개 신호등 그리드로 구분하였다. 이후 해당 신호등 중 스마트 횡단보도의 수요가 높은 지역을 판단하고자 입지지수라는 지표를 새로 정의하여 사용하였다. 이를 위해 어린이 교통사고 예측모형을 제시하고, 이에 필요한 주요 변수들을 선정한 뒤 각 변수의 가중치를 산출했다. 머신러닝 예측모형을 통해 도출한 창원시 맞춤형 스마트 횡단보도 입지지수로 총 103개의 스마트 횡단보도 최적입지를 선정했다. 이후, k-means 클러스터링을 통해 선정된 최적입지를 3가지 특성으로 분류하였고, 각 클러스터에 적합한 맞춤형 횡단보도 시설물을 제안하였다.

본 연구는 다음과 같은 학술적, 정책적 시사점을 갖는다. 먼저

본 연구는 기존의 스마트 횡단보도 관련 연구의 주제를 다양화했다는 점에서 의의를 가진다. 최근 많은 지자체가 스마트 횡단보도를 도입하고 있음에도 불구하고 기존의 선행 연구는 스마트 횡단보도의 시설물 개발에 대한 주제로 한정되어 있었다. 이에 본 연구는 스마트 횡단보도의 최적입지 선정 방법을 제안함으로써 관련 연구의 폭을 넓히는데 기여했다.

본 연구는 학술적 시사점 외에도 다양한 정책적 시사점을 갖는다. 본 연구는 창원시의 행정 문제 해결에 기여할 수 있을 것으로 기대된다. 창원시의 보행자 및 교통약자에 대한 교통안전지수는 최하등급(E)으로 타 시도와 비교하면 심각하게 낮은 상황이다. 하지만 본 연구의 결과를 토대로 창원시에 스마트 횡단보도를 유형별로 도입한다면, 창원시 교통사고 관련 문제를 효과적으로 해결할 수 있을 것으로 기대된다. 또한, 어린이를 고려한 스마트 횡단보도의 설치는 향후 노인을 포함한 다양한 교통약자를 보호하는데 효과적일 뿐만 아니라, 창원시의 행정 서비스 만족도를 향상하는데 기여할 수 있을 것으로 기대된다. 따라서 기타 취약계층을 중심으로 하는 스마트 횡단보도 최적입지 선정에도 중요한 연구지표가 될 것으로 기대된다. 마지막으로 본 연구는 지자체가 스마트 횡단보도 도입에 필요한 행정 예산을 효율적으로 운용하는데 기여하였다. 최적입지 선정에 그치는 것이 아니라 선정된 입지를 유형화하여 각 유형에 적합한 시설물을 제안함으로써, 각 입지에 불필요한 시설물이 설치되는 것을 방지하고 스마트 횡단보도의 효과를 극대화할 수 있을 것으로 기대된다. 이와 더불어 보행자 보호를 위한 규정을 강화하고, 사고예방을 위한 도로 설계 및 교통안전시스템의 프로토콜이 수립된다면 보다 안전한 보행환경을 구축할 수 있을 것이다.

하지만 본 연구에는 다음과 같은 한계점 또한 존재한다. 첫째, 횡단보도 이용률은 각 횡단보도에서 발생하는 교통사고 건수에 영향을 미치는 중요한 지표이나 유동인구 데이터 확보가 어려워 횡단보도 간의 이용률 편차를 반영하지 못했다. 추후 창원시 내 유동인구 데이터를 확보한다면 핫스팟 분석기법 등을 활용하여 시간대 및 연령대별 유동인구의 공간적 분포를 보다 정확하게 파악, 이를 반영하여 보다 정밀한 분석이 가능할 것으로 보여진다. 둘째, 어린이 교통사고 발생지점에 대한 좌표 데이터 확보가 가능하다면 보다 세밀한 교통사고 예측모형 설계가 가능할 것으로 보인다. 더 나아가 교통약자(예: 어린이, 노인, 장애인) 관련 교통사고 발생지점 좌표 데이터를 추가된다면 후속연구에서는 더욱 더 심도 있는 분석이 가능할 것으로 보인다. 셋째, 본 연구에서는 창원시를 중심으로 스마트 횡단보도 최적입지 선정을 진행하였다. 해당 모형은 추후 어린이 인구 비율이 높은 지자체를 중심으로 적용이 가능할 것으로 판단된다. 넷째, 본 논문에서 제시한 스마트 횡단보도는 어린이의 인지적 특성을 중심으로 스마트 횡단보도의 유형을 구분하였다. 차후 더 많은 교통약자의 특성을 반

영한 스마트 횡단보도 도입을 위해서는 노인층 및 스마트폰의 이용으로 시선이 아래로 고정된 젊은층(예: 스몸비족)을 고려한 횡단보도 유형 확립 및 지수 개선이 필요할 것으로 보인다.

## 감사의 글

이 연구는 한국외국어대학교 교내연구지원사업 지원에 의하여 이루어진 것임; 이 논문은 2020년 대한민국 교육부와 한국연구재단의 지원을 받아 수행된 연구임 (NRF-2021S1A5A2A03061484)

## References

- An, K. E., Lee, S. W., Jeong, Y. J., Lee, M. W., Lee, D. H., Seo, D. M. (2016). A Study of Object Detection in Crosswalk Area using Smart Crosswalk System, Proceedings of the Korean Information Science Society Conference, pp. 1,498–1,500.
- Baek, T. H., Son, S. K., Park, B. H. (2016). Modeling Traffic Accident Occurrence Involving Child Pedestrians at School Zone, Journal of Korean Society of Transportation, 34(6), pp. 489–498.
- Beak, T. H., Park, B. H. (2013). Developing Children Traffic Accident Models in School Zone : Focused on Cheongju, Proceedings of the KOR–KST Conference, 68, pp. 134–138.
- Choi, D. G., Park, G. H. (2018). Analysis of Dispatch Strongpoint for the Fire Accidents Based on Spatial Location–Allocation Model in the Chungnam Province, South Korea, Journal of KARG, 24(2), pp. 267–278.
- Choi, H. W., Ko, B. H., Won, D. W., Yeo, H. Y., Yun, W. (2021). The Optimal Location for a Smart Bus Stop through Public Data Analysis: The Case of Changwon–si, Journal of Global Business Research, 33(2), pp. 17–33.
- Gan, J. H., Nam, S. H., Seo, Y. J., Jeon, E. J., Lee, K. C. (2019). A Proposal for the Location of Wind Road and Green Space for the Mitigation of Urban Heat Island Effect in Seoul : Focusing on the proposal for heat island index using clustering and variable importance by random forest, Journal of Digital Convergence, 6(1), pp. 27–38.
- Géron, A. (2017). "Hands–on machine learning with scikit–learn & tensorflow.", 2nd, o'reilly.
- Jang, S. H., Cho, H. E., Jeong, J. W. (2019). Design and Implementation of A Smart Crosswalk System based on Monitoring Vehicle Speed using Deep Learning, Proceedings of the Korean Institute of Information and Commucation Sciences Conference, 23(2), pp. 522–525.
- Jang, S. H., Cho, H. E., Jeong, J. W. (2020). Design and Implementation of A Smart Crosswalk System based on Vehicle Detection and Speed Estimation using Deep Learning on Edge Devices, Journal of the Korea Institute of Information and Communication Engineering, 24(4), pp. 467–473.
- Jeong, J W., Zhou, K., Oh, S. K., Ryu, B. G. (2021). Design of RBFNN Classifier Using Elastic Net and Its Application, The Korean Institute of Electrical Engineers Conference, pp. 1767–1768.
- Jerome H. Friedman. (2001). Greedy Function Approximation: A Gradient Boosting Machine, The Annals of Statistics, 29(5), pp. 1189–1232.
- Kang, W. G., Lee, G. S., Yoon, Y. G. (2016). Factor Analysis of Child–Pedestrian Accidents in Seoul, The KOR–KST Conference, pp. 233–238.
- Ke, G., Meng, Q., Finley, T., Wang, T., Chen, W., Ma, W., Ye, Q., Liu, T. (2017). Lightgbm: A highly efficient gradient boosting decision tree, Advances in neural information processing systems, 30, pp. 3146–3154.
- Kim, H. B., Kim, S. G., (2006). A Site Selection of Public Facility Based on An Accessibility Theory & GIS Spatial Analysis Technologies, KSCE Journal of Civil Engineering 26(3), pp. 385–391.
- Kim, J. I., Chung, H. W., (2001). GIS Applications for Optimum Site Selection of Public Facility, Journal of the KAGIS 4(4), pp. 8–20.
- Kim, J. Y., Lee, H. S., Oh, J. S. (2020). Study on prediction of ship's power using light GBM and XGBoost, Journal of Advanced Marine Engineering and Technology, 44(2), pp. 174–180.
- Kim, S. U., Bang, J. I., Hong, S. E., Kim, H. J. (2019). "A Study on the Prediction of Missing Value in Data of Air Pollution Using LightGBM", Korea Institute of Communication Science Conference, pp. 1029–1030.
- LightGBM. (2020). Parameters Tuning, <https://lightgbm.readthedocs.io/en/latest/Parameters-Tuning.html> (Dec. 30. 2021).
- Lee, D. H., Lim, H. S. (2021). Detection of inappropriate advertising content on SNS using k–means clustering technique, Proceedings of the Korea Information Processing Society Conference, pp. 570–573.

- Lee, G. W., Kim, T. H. (2019). Examining the Characteristics of Traffic Accidents Involving Elderly Drivers in Seoul, South Korea, Ministry of Land, Infrastructure and Transport 102(1), pp. 19–34.
- Lee, H. S., Kim, H. C., Lee, G. W. (2017). A Study on Priority Determination of the Local Logistics Policies in Chungcheongnam-do Province, Journal of Korea Planning Association, 52(7), pp. 109–120.
- Lee, H. J. (2020). After installing the smart crosswalk, traffic stop line violations decreased by 70%, <https://www.lak.co.kr/m/news/view.php?id=8372> (Dec. 11, 2021).
- Lee, H. J., Lee, S. G. (2021). Comparative Analysis of Machine Learning Models for the Prediction of Pedestrian Crash Severity: Focused on Balancing Pedestrian Crash Dataset, journal of Korean Society for Geospatial Information Science, 29(2), pp. 3–15.
- Lee, S. H., Woo, Y. H. (2018). A Study on the Improvement of Prediction Accuracy for Traffic Accident Models Using Machine Learning (Generalized Regression Neural Network), International Journal of Highway Engineering, 20(6), pp. 179–189.
- Lee, S., Lee, W. J. (2016). Development of a system for predicting photovoltaic power generation and detecting defects using machine learning, KIPS Transactions on Computer and Communication Systems, 5(10), pp. 353–360.
- Ministry of Land, Infrastructure and Transport. (2020). standard analysis model, <http://www.geobigdata.go.kr/portal/> (Aug. 20, 2021).
- Na, H. H., Kang, J. H., Kim, D. G. (2021). Deriving the function of a sign for a protected area using the KANO model, Proceedings of the KOR-KST Conference, 85(), pp. 186–186.
- Oh, S. H., Baek, J. H., Kang, U. G. (2021). Classification models for chemotherapy recommendation using LGBM for the patients with colorectal cancer, Journal of The Korea Society of Computer and Information, 26(7), pp. 9–17.
- Oh, S. T. (2021). Jeollanam-do, Pedestrian-tailored traffic safety policy has an effect, [http://m.sisatotalnews.com/article.php?aid=1632627152133515011#\\_enliple](http://m.sisatotalnews.com/article.php?aid=1632627152133515011#_enliple) (Nov. 29, 2021).
- Park, H. G. (2021). "96% of drivers ignored pedestrian crossings without signs", <https://m.mk.co.kr/uberin/read.php?sc=30000001&year=2021&no=1086938> (Dec. 29, 2021).
- Park, S. N., Lim, J. B., Kim, H. K., Lee, S. B. (2017). Accidents involving Children in School Zones Study to identify the key influencing factors, International Journal of Highway Engineering, 19(2), pp. 167–174.
- Pejman, A., Bidhendi, G. N., Ardestani, M., Saeedi, M., Baghvand, A. (2017). Fractionation of heavy metals in sediments and assessment of their availability risk: A case study in the northwestern of Persian Gulf, Marine pollution bulletin, 114(2), pp. 881–887.
- Pintel. (2021). Pedestrian crossing safety system, <http://www.pintel.co.kr/html/index/> (Jan. 11, 2022).
- Rousseeuw, P. J. (1987). Silhouettes: a graphical aid to the interpretation and validation of cluster analysis, Journal of computational and applied mathematics, 20, pp. 53–65.
- Seoul. (2016). [Smart Crosswalk] Service, <https://smart.seoul.go.kr/index.do> (Dec. 20, 2021).
- Seongdong-gu. (2020). Big data analysis for smart crosswalk target selection, <https://www.sd.go.kr/main/index.do> (Sep. 21, 2021).
- Song, H. G., Lee, G. H. (2017). Optimal Location Modeling of Early Voting Polls Considering Spatial Accessibility: Cases of Seocho and Gangnam-gu in Seoul, Journal of the Association of Korean Geographers, 52(6), pp. 827–843.
- TAAS. (2018). 2017 Traffic Accident Statistical Analysis Report, <http://taas.koroad.or.kr/> (Dec. 20, 2021).
- TAAS. (2019). 2018 Traffic Accident Statistical Analysis Report, <http://taas.koroad.or.kr/> (Dec. 15, 2021).
- TAAS. (2020). 2019 Traffic Accident Statistical Analysis Report, <http://taas.koroad.or.kr/> (Dec. 17, 2021).
- TAAS. (2021). 2019 Traffic Accident Statistical Analysis Report, <http://taas.koroad.or.kr/> (Dec. 21, 2021).
- Yeom, M., Kim, K. M. (2011). "Deriving the causes of low fertility and policy demand through cluster analysis.", Journal of Economics Studies, 29(1), pp. 163–190.
- Yun, H. Y., Koo, Y. S., Choi, D. R. (2017). "A Development of Ensemble Model Based on Cluster Analysis to improve PM10 Forecasting Accuracy: Focus on the Weighted Average Ensemble by Weather Cluster.", Journal of Korean Society of Urban Environment, 17(1), pp. 33–42.
- Yoon, B. J., Ko, E. H., Yang, S. R. (2016). The Study on Traffic Accident of Commercial Vehicle using Odered Logit Model, Proceedings of the Korean Society of Disaster Information Conference, pp. 265–266.
- Yoon, H. S., Shim, J. Y. (2021). Coffee shop location prediction study using machine learning : Focusing on Coffee Bean case, KMIS International Conference, pp. 183–188.