pISSN : 2508-3384 eISSN : 2508-3392
https://doi.org/10.22640/lxsiri.2021.51.2.83

Journal of Cadastre & Land InformatiX Vol.51 No.2 (2021)
한국국토정보공사「지적과 국토정보」제51권 제2호 2021년 12월 p. 83-106

머신러닝 기반 고속도로 내 수소충전소 최적입지 선정 연구*

A Study on the Optimal Location Selection for Hydrogen Refueling Stations on a Highway using Machine Learning

조재혁** · 김성수***
Jo, Jae-Hyeok · Kim, Sungsu

Abstract

Interests in clean fuels have been soaring because of environmental problems such as air pollution and global warming. Unlike fossil fuels, hydrogen obtains public attention as a eco-friendly energy source because it releases only water when burned. Various policy efforts have been made to establish a hydrogen based transportation network. The station that supplies hydrogen to hydrogen-powered trucks is essential for building the hydrogen based logistics system. Thus, determining the optimal location of refueling stations is an important topic in the network. Although previous studies have mostly applied optimization based methodologies, this paper adopts machine learning to review spatial attributes of candidate locations in selecting the optimal position of the refueling stations. Machine learning shows outstanding performance in various fields. However, it has not yet applied to an optimal location selection problem of hydrogen refueling stations. Therefore, several machine learning models are applied and compared in performance by setting variables relevant to the location of highway rest areas and random points on a highway. The results show that Random Forest model is superior in terms of F1-score. We believe that this work can be a starting point to utilize machine learning based methods as the preliminary review for the optimal sites of the stations before the optimization applies.

Keywords: Machine Learning, Hydrogen Logistics Network, Hydrogen Refueling Station, Optimal Location Selection, Preliminary Review Methodology

1. 서론

미세먼지로 인한 대기오염, 온실가스 배출에 따른 지구온난화 등 세계적으로 환경 위기의 심각성이 대두되면서 기존의 화석연료를 대체하고자 청정에너지원에 대한 관심이 증대되고 있다. 특히 온실가스는 이

* 이 논문은 2021년 대한민국 교육부와 한국연구재단의 인문사회분야 신진연구자지원사업의 지원을 받아 수행된 연구임 (NRF-2021S1A5A8073366)

** 경북대학교 경영학부 박사과정 School of Business Administration, Kyungpook National University (first author: jjg7578@knu.ac.kr)

*** 경북대학교 경영학부 부교수 School of Business Administration, Kyungpook National University (corresponding author: sungsukim@knu.ac.kr)

상기후 현상의 주요 원인으로 지목되었으며, 이에 세계 각국은 2015년 파리 기후변화협약 당사국총회에서 온실가스 감축 및 감축목표 수립을 합의하였다(외교부 2015). 또한, 신재생에너지와 같은 청정에너지의 활용은 기후변화 문제에 적극적으로 대응할 수 있으므로, 온실가스 감축을 위한 국제적인 노력에 힘입어 2050년까지 전 세계 에너지 소비량 중 약 60%가 신재생에너지로 조달될 것으로 전망되고 있다(Gielen et al. 2018).

이러한 신재생에너지 중에서 수소는 현재 가장 주목받는 에너지원으로 기존의 화석연료와는 달리 연소 시 수분 이외에 부산물이 발생하지 않아 청정에너지로 분류된다. 이에 따라 수소는 전력생산, 교통과 물류 등 현존하는 에너지 체계 내의 모든 부문에서 널리 적용될 것으로 예상되며, 탄소배출 저감을 달성하는 동력원으로써 기술 및 경제적으로 활용 가능한 대체 에너지가 될 것으로 기대된다(Brandon and Kurban 2017). 해외에서는 선진국인 미국, 유럽, 일본 등의 국가들이 정부 차원에서 수소 산업 육성과 관련 기술 개발을 정책적으로 지원하고 있으며, 민간 차원에서도 이에 발맞추어 적극적으로 대규모 투자 계획을 수립하는 추세이다(Li et al. 2019). 국내의 경우 정부는 미래에 수소경제를 선도하고자 한국이 경쟁 우위가 있는 수소자동차 및 수소연료전지를 양대 축으로 산업 생태계를 조성하기 위해 “수소경제 활성화 로드맵”을 발표하여 수소차 보급과 수소충전소 설치를 천명하였다(김정화 2020). 이처럼 수소는 국내외에서 주목받는 청정에너지원으로 장차 그 활용 분야가 확대될 것으로 보인다.

물류 분야에서도 수소를 기반으로 하는 물류 네트워크의 구축을 위한 노력이 이어지고 있다. 이는 교통 부문이 전 세계 이산화탄소 배출의 상당량을 차지할 뿐만 아니라 화물차를 비롯한 대형 차량이 교통 부문 온실가스 배출의 약 40% 비중을 차지하는 것에서 기인한다(Kluschke et al. 2019). 이러한 물류 네트워크에서 수소충전소 인프라는 수소추진 자동차의 보급을 지원하고 수소 물류 체계를 지탱하는데 중요한 역할을 수행한다(Alazemi and Andrews 2015). 수소충전소 인프라 관련 분야에서 중요하게 다뤄지는 주제 중 하나는 “수소충전소 최적 입지 결정 문제”이다. 이 주제는 기본적으로 시설 입지 의사결정(Facility Location Decisions)에 해당하는데, 충전소 시설이 충분히 확보되지 않거나 적절한 위치에 설치되지 않으면 대체 연료 자동차가 보급되는데 차질을 빚을 수 있다(Hosseini and Mirhassani 2015). 수소충전소의 최적 입지 문제를 다루는 연구들은 주로 수리적 모형에 기반한 최적화 기법을 방법론으로 활용하였으며 수소충전소와 관련된 다양한 변수들을 고려하여 의미 있는 연구성과를 이루어냈다(Lin et al. 2020). 이것은 시설 입지 문제가 전통적으로 최적화 분야에서 연구되어온 주제이기에 수소충전소 입지를 결정하는데도 이와 같은 최적화 기법들이 유용하게 사용될 수 있다는 점에 기인한다. 그러나 이러한 최적화 기법에 기반한 연구들의 경우 최적입지의 후보지들에 대한 공간적 특성을 의미하는 지리적 변수들을 충분히 고찰하지 않았다는 단점을 가지고 있어(Dagdougui 2012), 이러한 공간적 특성들을 고려하기 위한 방법론의 필요성이 제기된다.

본 연구는 이를 위해 머신러닝(Machine Learning) 모형들이 수소충전소의 최적입지를 결정하는데 있어 최적화 기법 적용 이전에 충전소 후보지의 공간적 특성에 대한 예비적 검토 방법론으로 활용할 수 있는지에 그 가능성을 확인하고자 한다. 머신러닝은 컴퓨터와 같은 기계를 대규모의 데이터로 학습시키고 그 데이터에서 특징, 패턴 등을 추출하여 의사결정 등에 유용한 정보를 얻는 것을 목적으로 하는 인공지능(Artificial Intelligence)의 한 분야이다(김은미 외 2020). 머신러닝 모형들은 현재 다양한 분야에서 그 성능의 우수성이 검증되었으나, 아직 고속도로 내 수소충전소 입지와 관련된 연구에서는 그 방법론으로

채택된 바가 없다. 물론 국내에 특정 도시 지역의 수소자동차 충전수요에 대응하기 위해 충전소 위치를 결정하는데 머신러닝을 활용한 연구들이 있으나(김수환·류준형 2020), 고속도로를 대상으로 한 연구는 없는 것으로 나타났다. 이는 수소추진 화물차가 물류 운송에 주로 주행할 고속도로에 대해서는 머신러닝을 활용한 연구가 부족함을 시사한다. 또한, 전술한 바와 같이 수소충전소 최적 입지를 결정하는 연구에서 가장 많이 활용된 최적화 기법이 충전소 입지 후보지의 공간적 특성을 충분히 고려하지 못하는 것에 반해, 머신러닝의 경우 대량의 데이터와 다양한 변수들을 사용하여 예측을 실시하는 특징으로 후보지들의 공간적 특성을 학습 데이터와 변수 선정에 충분히 반영할 수 있다는 점에서 머신러닝 모형의 채택은 타당하다고 볼 수 있다.

선행연구의 검토 결과 기존 연구들이 수소충전소 최적 입지 결정에 있어 최적화 기법만 사용하였음을 고려할 때 최적화 기법의 투입 변수 중 하나인 충전소 후보지의 공간적 특성을 1차적으로 검토하는데 머신러닝이 활용할 수 있을 것으로 사료된다. 그러므로 본 연구는 고속도로상의 수소충전소 최적 입지를 결정함에 있어 머신러닝이라는 방법론을 적용하였다는 점과, 추후의 연구에서 전통적인 최적화 기법만을 적용하였을 경우와 비교하여 후속 연구에 더욱 많은 공간적 특성을 반영하는 초석이 될 수 있다는 점에서 그 학술적 의의가 있다.

이때 머신러닝 모형들을 학습할 데이터를 구성하기 위해 고속도로 휴게소 및 고속도로상 무작위 지점들의 위치 관련 변수들을 독립변수로 사용하였으며, 종속변수로는 이진 정수(Binary Integer)의 값을 가지는 충전소 입지 여부로 선정하여 분류(Classification) 문제로 정의하였다. 여기서 고속도로 휴게소를 데이터에서 종속변수로 선정한 이유는, 현재 정부가 2019년 수소경제 활성화 로드맵의 후속조치로 발표한 “수소 인프라 및 충전소 구축 방안”에서 고속도로의 경우 고속도로 휴게소를 중심으로 수소충전소를 설치 및 운영할 예정이라는 정책 방향에 기반한 것이다. 로드맵에 따르면 정부는 장기적으로 고속도로에 2030년까지 75km 이내, 2040년까지 50km 이내에 수소추진 차량이 고속도로 내에서 이용 가능한 충전소가 설치될 수 있도록 할 계획을 가지고 있다(국토교통부 2019). 이에 따라 본 연구에서는 고속도로 휴게소와 고속도로상 무작위 지점들의 위치와 관련된 여러 변수를 독립변수로 고려하여, 실제 정부 정책을 바탕으로 충전소의 최적 입지를 결정하는데 있어 머신러닝의 활용 가능성을 확인하였다는 점에서 실무적 의의가 있다.

전술한 바와 같이, 본 연구에서는 머신러닝 모형들을 사용하여 최적 입지 문제를 접근하고자 한다. 본 연구에서 고려하고 있는 모형은 의사결정나무(DT, Decision Tree), 서포트벡터머신(SVM, Support Vector Machine), K-최근접이웃법(KNN, K-Nearest Neighbor), 랜덤포레스트(Random Forest), 서포트벡터머신 배깅 앙상블(SVM-Bagging Ensemble), K-최근접이웃법 배깅 앙상블(KNN-Bagging Ensemble), 보팅 앙상블(Voting Ensemble) 모형으로 총 7가지이다. 선행연구 검토 결과, 이들은 다양한 머신러닝 연구에서 사용되고 있을 뿐만 아니라, 분류 문제에 있어 탁월한 성과를 내는 것으로 나타났다. 이에 따라 본 연구는 이들 머신러닝 모형들을 고속도로상의 수소충전소 최적 입지 선정에 적용하여 모형 간의 성능을 비교하고자 한다.

본 논문은 다음과 같이 구성되어 있다. 제2장에서는 수소충전소 최적 입지 선정 문제의 선행연구, 머신러닝 모형 기반의 최적 입지 선정에 관한 선행연구들을 소개한다. 제3장은 데이터 수집과 전처리 방법, 선정된 예측변수 설명, 머신러닝 모형의 하이퍼파라미터 튜닝 등 제반 연구 절차에 대해 설명한다. 제4장은 각 모형의 구체적인 파라미터 값과 성능 평가척도를 소개하며 이에 기반한 모형들의 성과를 설명 및 기술한다. 제5장은 본 연구의 내용을 요약·정리한 후 연구의

시사점과 한계점 및 향후 연구 방향을 제시한다.

2. 이론적 배경

2.1. 최적화 기반 수소충전소 입지 선정 연구

수소충전소의 최적 입지를 결정하는 연구들은 주로 수리적 모형에 기반한 최적화 기법을 중심으로 수행되었다. 이러한 연구들은 주행 시간이나 충전시설로의 거리를 최소화하거나 혹은 충전 수요의 충족을 최대화하는 것을 그 목적으로 하며, 각 연구자의 관점에서 중요하게 생각하는 요소들을 최적화 모형의 투입변수로 고려하여 충전소의 최적 입지를 선정하고자 하였다(Ko et al. 2017). 수소충전소의 최적 입지를 결정하기 위한 모형은 상당수 존재하는데, 대표적으로 Set Covering Model, Maximal Covering Location Model, P-median, P-center, FCLM(Flow Capturing Location Model), FRLM(Flow Refueling Location Model) 등이 있다(Lin et al. 2020). 이들 모형은 충전 수요를 점 수요(Point Demand) 혹은 흐름 수요(Flow Demand)로 추정하는지에 따라 구분할 수 있다.

먼저 점 수요를 가정하는 모형으로, Set Covering Model은 사전에 정의된 거리 내에서 모든 수요를 충족하는데 필요한 최소한의 시설 수와 위치를 결정하는 모형으로, 많은 연구에서 수소를 비롯한 다른 대체연료 자동차의 충전시설 수와 위치를 결정하는데 사용되었다(Jeong 2016). Maximal Covering Location Model은 시설의 수가 주어져 있을 때 수요를 최대한 만족할 수 있는 시설의 위치를 결정하는 모형으로 주어진 시설 수의 제약하에서 수소충전 수요를 최대한 만족할 수 있는 위치를 결정하는 연구에 활용되었다(Frade et al. 2011). P-median은 시설과 수요지점 간의 거리를 최소화하는 것을 목적으로 하는 모형으로 수소 충전 시설과 충전수요지점과의 거리를 최소화하는 충전소 위치를 결정하는 연구에서 사용되었다(Muratori et al. 2018). P-center는 시설과 수요지점 간의 최대 거리를 줄이는 것을 목적으로 하는 모형으로 P-median과 더불어 최적 입지 연구에서 자주 사용된다(Elloumi et al. 2004).

전술한 모형들이 충전 수요를 점 수요(Point Demand)로 추정하는 것과 달리, FCLM과 FRLM은 그 수요를 교통흐름(Traffic Flow) 등으로 추정하는 방식을 사용한다(Lin et al. 2020). FCLM와 FRLM은 차량이 주행하는 경로를 완주할 수 있도록 차량의 항속가능거리 내에서 충전이 가능한 지점에 충전소 위치를 결정하여 최대한의 교통흐름 수요를 만족할 수 있도록 하는 모형으로써 수소충전소의 최적 입지 결정 문제에서도 활용된 바가 있다(Kuby et al. 2009). 두 모형의 차이는 FCLM이 특정 경로에서 1개의 충전시설이 경로 내의 충전 수요를 모두 충족할 수 있다고 가정하는 반면, FRLM은 장거리 주행 시 특정 경로에 복수의 충전시설이 있음을 가정한다는 점에서 FCLM보다 더욱 현실적이면서 우수한 모형으로 분류된다.

선행연구를 검토한 결과, 고속도로상 수소충전소의 위치를 결정하는 분야에서는 머신러닝 모형을 활용한 연구가 없는 것으로 나타났다. 전술한 바와 같이 최적화 모형은 시설 입지 연구들에서 그 연구방법론으로 주로 사용되어 왔고 많은 학문적·실무적인 발전을 이루었다. 이러한 연구 동향 속에서 본 연구가 최근 그 성과를 인정받고 있는 머신러닝 모형들을 활용하고자 하는 시도는 의미가 있을 것으로 생각된다.

2.2. 머신러닝 기반 최적입지 선정 연구

머신러닝은 인공지능 분야 중 하나로 대량의 데이터로 기계를 학습한 후 데이터에서 특징, 패턴 등을 추출하여 다양한 의사결정에 유용한 정보를 제공하는 기법이다. 머신러닝을 통해 다룰 수 있는 문제들은 분류(Classification), 회귀분석(Regression Analysis),

군집화(Clustering) 등 그 종류가 매우 다양하다(Sharma and Kumar 2017). 이 중 분류(Classification) 문제는 분석대상들의 종속변수 값을 특정 범주(Category)로 구분하는 것(Teng and Gong 2018)인데, 머신러닝 관점에서 시설 입지 문제는 시설의 입지 여부를 결정하는 것이므로 분류 문제로 정의할 수 있다.

머신러닝을 기반으로 최적 입지를 결정하는 연구로, 김수환·류준형(2020)은 서울특별시의 수소충전소 최적 위치를 결정하고자 머신러닝 모형인 K-중심자 군집화(K-medoids Clustering)를 사용하여 서울 내 주유소와 LPG충전소를 수소충전소의 잠재적인 위치로 가정한 상태에서 서울 내 각 자치구의 인구밀도, 등록 차량 수, 차량 종류 등을 데이터를 활용해 충전소의 위치를 예측하고자 하였다. 이우진·전준헌(2021)은 서울특별시의 전기 공공자전거 대여소의 최적입지를 예측하고자 비지도학습 머신러닝(Unsupervised Machine Learning) 모형인 K-means 군집분석을 사용하여 대여소의 최적 입지인 법정동을 선정하였는데, 최적 입지와 관련된 여러 요인에 대해 가중치 기반 분석을 실시하고 청년층과 장년층을 대상으로 서울 내 특정 자치구 내 대여소의 최적 입지를 예측하고자 하였다. Damavandi et al.(2019)은 최적의 소매점(Retail Store) 입지를 선정 및 각 입지들의 순위를 결정하는데 있어 머신러닝 모형을 사용하였으며, 많은 선행연구에서 분류 문제에 활용된 바 있는 다양한 머신러닝 모형들의 분석 결과를 비교함으로써 그 성능의 우수성을 검증하고 머신러닝 모형들이 최적 소매점 입지 결정에 활용될 수 있다고 말하였다. Zaheer et al.(2021)은 머신러닝 모형 중 하나인 심층신경망(Deep Neural Network) 모형을 사용하여 최적의 학교부지를 선정하는 연구를 수행한 결과 높은 정확도 성과를 보여 최적 입지 선정 문제에 있어 머신러닝 모형이 적용될 수 있음을 시사하였다.

선행연구의 검토 결과 머신러닝 모형은 많은 선행연구에서 밝혀진 성능의 우수성에도 불구하고 최적 입지를 결정하기 위한 문제에서는 활용된 사례가 매우 적었다. 또한, 대부분의 연구들은 최적 입지를 결정하는 연구의 경우 머신러닝 모형을 적용한 연구가 있으나 도시 등 특정 지역만을 대상으로 최적 입지를 결정하여 국가 단위의 지역을 대상으로 적용한 연구는 없는 것으로 나타났다. 그러므로 본 연구는 선행연구들과는 달리 국가 단위의 지역에 대해 머신러닝 모형을 적용한다는 점에서 그 의의가 있을 것으로 사료된다. 또한, 선행연구의 검토 결과 최적 입지를 결정하는 문제에 머신러닝 기법을 적용할 때 변수의 연구 목적에 대한 부합 정도 및 예측 대상과 독립변수 사이의 연관성 등이 예측에 사용할 변수의 선정이 중요한 요소인 것으로 나타났다. 그러므로 본 연구에서도 이러한 연구 목적과 예측 대상과의 관련 정도와 같은 사항들을 충분히 고려하여 연구를 수행하고자 하였다.

선행연구의 탐색 결과, 현재 수소충전소의 최적 입지를 결정하는 연구에 있어 머신러닝과 최적화 기법을 모두 활용한 연구는 없는 것으로 나타났다. 그러므로 수소충전소의 위치를 결정함에 있어 머신러닝을 통해 충전소 후보지에 대해 예비적 차원에서 검토를 수행하기 위한 본 연구는 최적화 기법을 적용하여 최적의 입지를 결정하기 위한 후속 연구에서 더욱 의미있는 결과를 도출할 수 있을 것으로 판단된다.

3. 연구설계

3.1. 데이터 수집 및 전처리 과정

본 연구에서 머신러닝 모형들의 학습 및 예측 성능 비교를 위하여 사용된 데이터 수집과정을 설명하면 다음과 같다.

3.1.1. 데이터 수집 및 전처리 과정

먼저 분석은 처음에 전국의 휴게소 296곳과 고속도로상의 랜덤 지점 600곳으로 총 840곳을 대상으로 하

Table 1. Decision Variables and Independent Variables

Variables	Name	Description
Dependent Variable	Candidate for a charging Station	Expressway service station(1) and random point on highway(0)
Independent Variable	Land price	Land price of Expressway service station areas and random points
	Expressway route	Route name the points are located in
	Distance from hydrogen sources	Distance between the points and by-product hydrogen sources
	Distance from tollgates	Distance between the points and tollgates on the expressway
	Distance from the 5 nearest points	Distance between each of the points and 5 nearest points on the highways

였다. 랜덤 지점은 머신러닝 모형들의 학습을 위한 데이터 수를 확보하기 위한 것으로, 고속도로 충전소 후보지인 휴게소와 동일한 도로망에 위치한 무작위 지점들을 함께 데이터를 구성하여 머신러닝 모형이 실제로 충전소 후보지 여부를 잘 예측할 수 있는지 확인하는 목적으로 데이터에 포함하였다. 또한, QGIS 3.16.9 프로그램을 사용할 경우 약 50,000곳의 고속도로상 랜덤 지점을 생성할 수 있으나, 이들 데이터를 모두 머신러닝의 학습 및 테스트용 데이터에 포함할 경우 특정 범주의 데이터가 지나치게 많아져 극심한 데이터 불균형성 문제를 초래할 수 있다. 이러한 불균형성은 머신러닝 모형의 예측력의 저하를 가져올 수 있으므로(Kaur at al, 2019), 50,000곳의 무작위 지점 중 일부인 600곳만을 무작위로 추출해 데이터에 포함하여 불균형성 문제를 줄이고자 하였다. 전국의 휴게소 정보는 정부가 운영 중인 공공데이터포털의 “전국휴게소정보표준데이터”를 활용하였는데, 본 연구가 고속도로를 주행하는 수소화물차를 위한 수소충전소의 최적입지를 결정하는 목적에 부합하기 위해 일반휴게소로 분류되지만 국도와 같은 비고속도로에 위치한 휴게소를 데이터에서 제외하였다. 또한, 고속도로에 위치하지만 임시로 운영 중인 임시휴게소도 휴게소 데이터에서 제외하였다. 고속도로상의 무작위 지점은 국가교통데이터베이스(KTDB)에서 제공하고 있는 전국 도로망데이터에서 GIS 프로그램 중 하나인 QGIS 3.16.9를 활용하여 무작위로 추출하였다. 그리고 QGIS 3.16.9프로그램을 통해 고속도로의 휴게소와 무

Figure 1. Location of Expressway Service Stations and Random Points

Table 2. Process of Selecting Dataset for Analysis

	Description
Initial dataset	- Hydrogen Refueling station candidate sites : 296 expressway service stations - None-candidate sites : 600 random locations on the highway
Data Preprocessing	- The positions in the below are excluded from the dataset. 1) Service stations which are not on the highway 2) Temporary service stations 3) the sites that land price data are not available
Final dataset	- 296 expressway service stations - 485 random locations on the highway

작위 지점들의 경도 및 위도 정보를 추가하여 각 지점에 대한 위치 관련 변수들의 값을 찾기 위한 기초 정보로 사용하였다. 이때, 고속도로상의 무작위 지점에 대해 후술할 독립변수 중 하나인 개별공시지가 정보가 존재하지 않을 때에는 해당 지점 데이터에 결측치(Missing Value)가 존재하는 것으로 간주하여 분석대상에서 제외하였다. 이에 따라 최종적으로 분석에 활용한 데이터로 고속도로휴게소 240곳과 무작위 지점 485곳을 활용하였다. 이러한 분석대상과 관련된 내용을 정리하면 Table 1과 같으며, 각 지점을 지도상에서 표현하면 Figure 1과 같다.

3.1.2. 독립변수 및 종속변수

본 연구에서 사용한 데이터를 구성할 때 선정한 독립변수 및 종속변수들의 정의와 각 지점의 변수들의 값을 추출하는 과정을 설명하면 다음과 같다. 그리고 데이터에 대한 정보를 요약하여 나타내면 다음의 Table 2와 같다. 선행연구에서 검토한 바와 같이, 머신러닝 모형을 최적 입지 문제에 적용할 때에는 독립변수의 선정에 있어 연구 목적과의 부합성, 종속변수와의 연관성 등을 종합적으로 고려할 필요가 있다. 본 연구는 고속도로상의 수소충전소 최적 입지를 결정하는 것을 목적으로 하며, 수소충전소 입지는 충전소 후보지에 대한 위치 관련 데이터들이 충전소 입지 여부와 관련성이 높으므로 위치와 관련된 변수들을 후술할 독립변수들로 선정하였다. 이러한 변수값 추출 및 전처리를 위해 주로 Python 3.8.5 프로그램을 사용하였으며, 그 외에 변수값 추출을 위해 GIS 프로그램 중 오픈 소스(Open Source)에 해당하는 QGIS 3.19.6 프로그램과 Google Map API를 사용하였다. 이때, 후술할 독립변수인 수소공급지와의 거리, 영업소와의 거리, 최단지점 5곳과의 거리는 전술한 국가교통데이터베이스에서 제공하는 전국도로망 데이터를 기반으로 거리지점 간에 자동차가 실제로 주행하는 도로상의 거리를 기준으로 거리를 측정하여 거리 관련 데이터의 타당성을 확보하였다.

1) 종속변수 및 변수값 추출과정

종속변수는 충전소 후보지 여부이다. 앞서 언급한 바와 같이 본 연구에서는 정부의 정책에 따라 고속도로 휴게소에 수소충전소를 설치한다고 가정하므로 고속도로 휴게소의 종속변수값을 1로 라벨링(labelling)하고 고속도로상의 랜덤 지점은 0으로 라벨링하여 변수값이 이진 정수를 가지도록 처리하였다. 이러한 라벨링 처리는 머신러닝 연구에서 분류 문제를 해결할 시 종속변수값 상에서 서로 다른 범주에 속하는 데이터들을 구분하기 위해 필수적인 과정인데, 본 연구에서 분석에 사용한 모형들은 그 학습 시 종속변수 값이 제시된 데이터를 학습하여 예측을 실시하는 지도학습(Supervised Learning) 모형이므로 고속도로 휴게소

여부를 구분하고자 각각 1과 0으로 종속변수값을 설정하였다.

2) 독립변수 및 변수값 추출과정

① 개별공시지가

개별공시지가는 국토교통부의 국토교통용어사전에 따르면 “국토교통부장관이 매년 공시하는 표준지공시지가를 기준으로 시장, 군수, 구청장이 조사한 개별토지의 특성과 비교표준지의 특성을 비교하여 토지가격비준표상의 토지특성차이에 따른 가격배율을 산출하고, 이를 표준지공시지가에 곱하여 지가를 산정한 후 감정평가업자의 검증을 받아 토지소유자 등의 의견수렴과 시, 군, 구 부동산가격공시위원회 심의 등의 절차를 거쳐 시장, 군수, 구청장이 결정, 공시하는 개별토지의 단위면적당 가격(원/제곱미터)”을 말한다. 즉, 개별공시지가는 국토교통부가 공시하는 표준공시지가에서 해당 토지의 속성에 따라 조정된 토지가격이다. 이러한 개별공시지가는 국가가 보증하며 과세, 토지거래 등의 기준이 되는 공적지가라고 할 수 있다(이창한·성춘자 2018). 설치 지역의 토지가격은 수소충전소의 설치 시 발생하는 비용 중 하나이므로 개별공시지가를 각 지점과 관련된 변수 중 하나로 선정하였다. 또한, 많은 선행연구에서 수소충전소 부지 매입비용을 최적 입지를 결정하기 위한 중요한 변수 중 하나로 다루고 있어(Li et al. 2019), 본 연구에서도 이를 반영하고자 개별공시지가를 독립변수에 포함하였다.

토지와 관련된 비용을 측정함에 있어 가장 타당한 측정치는 실제 토지의 매매가일 것이다. 그러나 실제 토지의 매매가의 경우 해당 토지의 매매거래일이 발생 시점으로부터 상당히 시간이 지난 경우가 대다수이기 때문에 해당 토지 매매가격을 변수값으로 사용하는 것은 적절하지 않다. 이와 같은 현실적인 이유를 반영하여 본 연구에서는 그 가격이 매년 갱신되는 개별공시지가를 각 지점의 토지 실제매매가격에 대한 대용치로 활용하였다. 지점별 개별공시지가는 정부공공데이터포털의 “국토교통부_개별공시지가정보서비스 API”를 통해 획득하였다.

② 고속도로 노선

고속도로 노선은 학습용 데이터를 구성하는 각 지점이 위치한 고속도로를 의미한다. 한국도로공사에 따르면 전국의 고속도로는 2021년 기준 현재 한국도로공사 고속도로 30개, 민자고속도로 16개로 총 46개 노선이 운영 중이며, 노선별 길이는 최장노선인 경부고속도로의 416.05km부터 최단노선인 남해제3지선 고속도로의 15.26km까지 매우 다양하다. 고속도로에는 휴게소들이 위치하고 있는데, 이러한 노선의 길이에 따라서 휴게소의 개수가 상이할 수 있다. 이는 고속도로 노선의 길이가 길어짐에 따라 수소충전소의 수 또한 증가할 수 있음을 의미하며, 본 연구가 고속도로 휴게소를 수소충전소의 잠재적인 후보지로 고려하고 있고 노선별 길이에 따라 존재하는 후보지의 수 또한 달라질 수 있어 노선을 독립변수로 타당하다는 판단하에 해당 변수를 포함하였다.

고속도로 노선은 고속도로 휴게소의 경우 “전국휴게소정보표준데이터”에서 제공하고 있어 해당 노선명을 사용하였다. 고속도로상의 무작위 지점은 국가교통데이터베이스의 도로망 데이터에서 제공하는 노선명을 사용하되, 두 데이터의 노선명 분류가 다른 경우에는 전국휴게소정보표준데이터의 노선명 분류로 통일하였다.

③ 수소공급지와의 거리

수소공급지와의 거리는 학습용 데이터의 각 지점과 전국 부생수소 생산지와의 거리를 의미한다. 우리나라는 수소경제 구축 단계에 있어 초기에 해당하므로 국내의 현존하는 수소인프라를 활용하면서 수소자동차가 보급될 수 있는 환경을 조성하는 것이 필요하다(김봉진 외 2014). 우리나라는 울산석유화학단지를 포함한 여러 대규모 석유화학단지에서 석유화학제품

생산 시 그 부산물로 부생수소가 발생하며, 부생수소 발생량은 연간 5만 톤으로 이 양은 대략 25만대의 수소자동차 수소 수요를 만족할 수 있다(대한민국 정책브리핑 2020). 또한, 우리나라는 부생수소를 충전소의 외부에서 수급하여 충전하는 고정형 충전소 위주로 설치가 추진되고 있다(정기대 2019). 따라서 현재 수소자동차 보급 초기 단계에 해당하는 우리나라의 경우, 부생수소 생산지에서 수소를 공급받는 것으로 가정하고 지점별로 이러한 생산지와의 거리를 독립변수에 포함하는 것이 타당할 것으로 사료된다.

본 연구에서는 전국의 대규모 석유화학단지를 수소공급지로 설정하고, 각 지점과의 거리를 QGIS 3.19.6 프로그램을 이용하여 산출하였다. 이때, 각 지점과 수소공급지와의 거리는 실제 도로망에 따라 이동하는 최단거리로 설정하여 현실성 있는 거리값이 될 수 있도록 하였다.

④ 최단지점 5곳과의 거리

최단 지점 5곳과의 거리는 특정 지점에서 최단거리에 위치한 지점 5곳과의 실제 도로상 거리를 말한다. 화석연료의 대체연료로 주행하는 자동차의 경우 대부분 그 보급이 초기 단계이고 운영 중인 충전소의 수가 제한적이라는 사실로 인해 연료의 충전을 위해 상당히 긴 거리를 주행해야 하는 특성을 가지고 있으므로(Lin et al. 2018), 충전소 후보지 간의 공간적 거리는 중요하게 고려해야 할 요소라고 할 수 있다. 이러한 맥락에서 지점별로 상대적으로 많은 수인 5곳의 최단지점을 고려하여 머신러닝을 학습하기 위한 데이터에 각 지점 간의 거리적 분포를 폭넓게 반영하고자 하였다. 자동차가 영업소에서 출발하여 특정 경로를 따라 주행 시 수소의 충전이 필요한 경우가 존재할 수 있다. 이때, 충전소 간의 거리가 너무 멀 경우 자동차가 주행 도중 연료가 소진될 수 있는 문제가 발생할 수 있으므로 각 지점 간의 거리도 고려해야 할 것이다. 본 연구에서는 지점별로 최단거리에 위치한 지점 5곳의 거리를 변수로 사용하였다. 이는 특정 고속도로의 영업소에서 자동차가 출발하였지만 도착지는 다른 노선일 수 있으며, 특정 지점에서 수소를 충전하고 난 후 다시 주행 시 다음 충전소와의 거리 또한 너무 멀어서는 안 된다는 점을 고려한 것이다.

본 연구에서는 각 지점별로 최단 지점 5곳과의 거리를 산출하기 위해 QGIS 3.16.9 프로그램에서 학습용 데이터에 포함된 모든 지점 간의 최단경로 거리를 그 값으로 하는 기종점쌍 거리 행렬(OD Distance Matirx)을 생성하였다. 그 후 해당 행렬을 대상으로 지점별로 다른 지점과의 거리값이 가장 작은 5곳을 추출하였다.

⑤ 영업소(Tollgate)와의 거리

영업소와의 거리는 각 지점과 지점들이 위치한 고속도로 영업소와의 거리를 말한다. 영업소는 자동차가 고속도로에서 주행 시 출발지(Origin)와 도착지(Destination)의 역할을 하는데, 자동차가 고속도로를 통해 특정 경로를 주행할 때 전체 주행거리를 결정하는 기준은 실질적으로 영업소라고 할 수 있다. 이때 출발지 혹은 도착지에 해당하는 영업소에서 수소충전소까지의 거리가 자동차 항속가능거리보다 멀다면 자동차는 특정 경로의 주행 중에 연료가 소진되는 문제가 발생할 수 있다. 이러한 상황은 최적화 기반의 수소충전소 최적 입지 연구에서도 고려된 바 있는데, 김거중 외(2019)는 FRLM의 응용 모형을 제시하면서 고속도로상의 특정 수소충전소 조합이 자동차의 주행가능여부를 결정할 시 자동차의 출발지와 도착지를 영업소를 기준으로 설정하여 자동차의 특정 경로 주행 가능여부를 판단하였다.

이에 따라 본 연구에서는 고속도로휴게소와 영업소 간의 거리를 독립변수 중 하나로 채택함으로써, 고속도로 주행의 출발지와 도착지가 되는 영업소와 고속도로상 특정 지점과의 거리가 적정한 수준을 유지해야 하는 점을 반영하고자 하였다. 영업소와의 거리는

Figure 2. Correlation Matrix Plot

수소공급지와의 거리와 동일하게 QGIS 3.19.6 프로그램을 사용하여 영업소와 고속도로상의 특정 지점 간에 거리를 실제 주행도로에 기반하여 산출하였다.

본 연구에서 선정한 독립변수 및 종속변수 간의 상관관계를 파악하고자 변수 간의 상관관계를 분석 및 시각화한 결과는 Figure 2에 제시되어 있다. 분석 결과 앞서 제시한 변수들 범주 간에 그 상관계수가 ±0.8을 초과하는 변수는 존재하지 않았다. 최단지점 5곳과의 거리에 해당하는 x5-x8 사이에서는 상대적으로 높은 상관계수가 관찰되었지만, 본 연구에서는 최단지점 5곳과의 거리를 하나의 독립변수군으로 가정하여 분석을 실시하므로 독립변수 선정의 타당성을 저해하지 않을 것으로 사료된다.

3.2. 모형 평가척도

앞서 설명한 변수들을 토대로 머신러닝 모형들을 학습하고 분석할 경우, 이러한 분석 결과들을 평가할 수 있는 척도가 필요하다. 본 연구에서는 고속도로 수소충전소의 최적입지 문제를 머신러닝 모형으로 해결하고자 하였는데 이는 이진 분류 문제(Binary Classification Problem)에 해당한다. 이진 분류 문제에서는 예측 결과에 대한 경우의 수가 0 또는 1의 두 가지이므로 다음의 Table 3에 제시된 혼돈행렬(Confusion Matrix)로 가능한 모든 경우를 분류할 수 있다.

Table 3. Confusion Matrix

Actual \ Predict	Positive	Negative
Positive	True Positive (TP)	False Negative (FN)
Negative	False Positive (FP)	True Negative (TN)

본 연구에 따라 혼돈행렬의 구성요소를 설명하면 다음과 같다. TP는 실제 고속도로 휴게소를 머신러닝 모형이 고속도로로 예측한 경우이고, FP는 실제로는 고속도로상의 무작위 지점이지만 머신러닝 모형이 고속도로로 예측한 경우를 말한다. TN는 실제 고속도로상의 무작위 지점을 머신러닝 모형이 고속도로상의 무작위 지점으로 예측한 것을 말한다. FN는 실제 고속도로 휴게소를 모형이 고속도로상의 무작위 지점으로 예측한 것을 말한다.

혼돈행렬은 모형들의 예측 성과를 측정하는 여러 평가척도들을 산출하는데 활용할 수 있다. 머신러닝 모형을 평가하는 척도의 종류는 다양하지만, 대표적으로 정확도(Accuracy), 정밀도(Precision), 재현율(Recall), F1-점수(F1-Score)가 머신러닝 관련 연구에서 활용된다. 이러한 혼돈행렬을 바탕으로 4가지 평가척도들의 산식을 표현하면 다음의 Table 4와 같다.

머신러닝의 성과를 평가하는데 있어 전체 데이터에서 정확하게 예측한 것의 비율을 나타내는 정확도가 가장 직관적으로 이해하기 쉬운 척도이지만, 보수적인 관점에서는 정밀도와 재현율의 기하평균인 F1-점수를 머신러닝 모형의 평가척도로 활용하는 것이 바람직하다(이용준·선종완 2020). 그러므로 본 연구에서는 4가지 평가척도 결과를 모두 제시하지만, 모형들

Table 4. Evaluation Metrics

Evaluation metrics	Formulation
Accuracy	$\frac{TP+TN}{TP+TN+FP+FN}$
Precision	$\frac{TP}{TP+FP}$
Recall)	$\frac{TP}{TP+FN}$
F1-score	$\frac{2(Recall \times Precision)}{Recall+Precision}$

의 분석 결과에 대한 최종 평가는 F1-점수에 따라 실시한다.

3.3. 분석을 위한 머신러닝 모형

본 연구에서 분석을 위해 사용된 머신러닝 모형들은 단일 분류기 모형과 앙상블 모형들로 구성된다. 단일 분류기 모형은 의사결정나무(Decision Tree), K-최근접이웃법(K-Nearest Neighborhood), 서포트벡터머신(Support Vector Machine) 총 3가지이며, 앙상블 모형은 배깅 앙상블(Bagging Ensemble)과 보팅 앙상블(Voting Ensemble) 모형으로 2가지이다. 이러한 모형들을 설명하면 다음과 같다.

3.3.1. 의사결정나무(Decision Tree)

의사결정나무는 어떤 문제의 해결을 위한 의사결정 규칙을 나무 구조로 표현하여 목표를 분류 및 예측해 나가는 지도학습(Supervised Learning) 모형 중 하나이다(정용찬 외 2021). 의사결정나무는 활용가능한 학습 데이터를 여러 부분집합으로 나눈 후 이들 집합들을 사용하여 모형들을 학습하는데, 뿌리 노드(Root Node)에서 전체 데이터가 점차 작은 단위의 노드들로 나뉘어 마지막 시점의 노드인 잎 노드(Leaf Node)로 분기되는 과정을 거쳐 이들 노드들을 대상으로 학습 및 예측을 실시하는 머신러닝 모형이다(Friedl and Brodley 1997). 이러한 의사결정나무 모형은 분류 문제에 있어 타 머신러닝 모형에 비해 비슷하거나 더 우수한 예측 결과를 보이는 것으로 나타났다(Priyam et al. 2013). 또한, 의사결정나무는 상대적으로 단순하면서도 적용하기 쉬우며 통계학, 데이터 마이닝 등 다양한 분야에서도 자주 활용되는 예측 기법이다(Sharma and Kumar 2016).

3.3.2. 서포트벡터머신(Support Vector Machine)

서포트벡터머신은 복수의 범주(Category)를 가지는 학습용 데이터를 통해 모형을 학습한 후 이를 토대로 데이터들을 가장 잘 분류하는 최적의 초평면(Hyper Plane)을 생성하여 추후 새로운 데이터가 입력되었을 때 해당 데이터의 범주를 분류하는 지도학습 모형 중 하나이다(Pradhan 2012). 즉, 초평면에서 가장 가까운 벡터들 간의 거리에 해당하는 마진(Margin)이 최대가 되도록 하여 데이터의 집합을 최적으로 분할되는 최대 마진의 평면을 찾아 데이터들을 분류한다(박종훈 외 2006; 박지만 외 2018). 서포트벡터머신은 분류 정확도가 높고 모형의 유연성이 높아 분류 문제를 해결하는데 있어 널리 활용되고 있으며(김은경 외 2016), 특히 이진 분류 문제(Binary Classification Problem)에서 그 분류 능력이 탁월하여 다양한 분야에서 범주를 구분하는 문제를 해결하기 위한 방법으로 자주 사용되는 것으로 나타났다(Cervantes et al. 2020).

3.3.3. K-최근접이웃법(K-Nearest Neighbor)

K-최근접이웃법은 주어진 학습용 데이터를 바탕으로 모형을 학습한 후 새로운 데이터가 주어졌을 때 해당 데이터와 가장 근접한 데이터에서 빈도(Frequency)가 높은 범주로 새로운 데이터가 포함될 범주를 결정 혹은 예측하는 머신러닝 모형이라고 할 수 있

다(김남진·배영철 2018). 이때 가장 근접한 데이터를 선정하는데 있어 K-최근접이웃법은 유클리디언(Euclidean) 거리, 민코스키(Minkowsky) 거리, 맨하탄(Manhattan) 거리 함수 등 다양한 거리 함수(Distance Function)를 사용할 수 있는데 대표적으로 유클리디언 거리 함수가 K-최근접이웃법을 활용한 연구에서 데이터를 분류하기 위한 함수로 많이 활용된다(Hu et al. 2016). K-최근접이웃법은 모형의 특성상 복잡한 연산을 필요로 하지 않는 등 매우 간단하고 적용하기 쉬우며 분류 문제를 해결하는데 있어서도 널리 활용되는 머신러닝 모형 중 하나이다(Gazalba et al. 2017). 또한, 이미지 분류 등에 있어서도 K-최근접이웃법은 그 유용성을 보이는 등 많은 분야에서 활용되고 있다(Kim et al. 2012).

3.3.4. 배깅 앙상블(Bagging Ensemble)

앞서 소개한 머신러닝 모형들은 단 하나의 모형으로 분석하는 단일 분류기(Classifier) 모형이다. 이러한 단일 분류기에 비해 더욱 좋은 성과를 얻는 것을 목적으로 복수의 분류기를 결합하여 사용하는 앙상블(Ensemble) 모형도 다양한 연구에서 사용되기 시작하였다. 앙상블 모형은 일반적으로 단일 분류기의 일반화 성능을 증진할 뿐만 아니라 앙상블 분류기를 구성할 때 그 성과가 더욱 좋은 것으로 나타났다(민성환 2016).

먼저 배깅 앙상블은 분류 혹은 추정 문제 등에 있어 불안정성을 개선하고자 개발된 것으로, 부트스트랩(Boostrap)을 사용하여 학습용 데이터에서 표본들을 복원추출하고 해당 표본들로 기저 분류기를 학습한 결과들을 평균 혹은 다수결을 통해 모형의 최종 결과를 결정하는 앙상블 기법이다(Bühlmann 2012). 이때, 배깅 앙상블 모형은 부트스트랩을 통해 생성된 각 표본에 대해 동일한 종류의 기저분류기를 적용하여 학습 및 예측이 이루어지는 특징을 가지므로 데이터에 대한 변화를 주는 앙상블 모형이다(Shah et al. 2015). 배깅 앙상블을 활용해 분류 문제를 해결하는 경우 모형에 대해 적절한 하이퍼 파라미터(Hyper Parameter)를 설정할 경우 단일 분류기의 성과보다 그 성능이 더욱 우수한 것으로 나타났다(Kim et al. 2002). 그리고 많은 선행연구에서 배깅 앙상블 모형은 분류 문제에 있어 높은 분류 정확도와 낮은 FP비율을 제공하는 것으로 나타났으며 짧은 모형 구성 시간 등 많은 장점을 가지고 있는 모형임을 입증하였다(Gaikwad and Thool 2015).

3.3.5. 보팅 앙상블(Voting Ensemble)

보팅 앙상블은 2개 이상의 서로 다른 기저 분류기들을 결합하는 앙상블 모형으로 분류기들의 분석 결과들을 투표(Voting)를 통해 최적의 분류 및 예측 결과를 결정하는 기법이라고 할 수 있다(김한민 2020). 보팅 앙상블은 서로 다른 기저 분류기들을 사용하여 기저 분류기에 변화를 준다는 점에서 전술한 배깅 앙상블과 차이점이 존재한다. 보팅 앙상블은 각 기저 분류기의 분석 결과들에 대해 다수결 투표를 진행하여 가장 많이 출력된 결과가 앙상블 모형의 최종 결과가 되는 하드 보팅(Hard Voting)과 각 분류기들이 예측한 확률 및 가중치를 이용해 최종 확률을 계산 후 이를 반영해 최종 결과를 결정하는 소프트 보팅(Soft Voting)으로 구분할 수 있다(전병욱 외 2021). 많은 선행연구에서 단일 분류기만을 사용하였을 때와 비교하여 보팅 앙상블을 통해 다양한 종류의 분류기를 결합한후 최종 결과를 결정한다면 분류 문제에서 더욱 우수한 결과를 산출하는 것이 확인되었다(Gandhi and Pandey 2015).

본 연구에서는 고속도로 내 수소충전소의 최적 입지를 결정하는데 있어 앞서 소개한 머신러닝 모형들을 사용하여 그 적용가능성을 확인하고자 한다. 구체적으로 본 연구는 의사결정나무, 서포트벡터머신, K-최근접이웃법이라는 3가지 단일 분류기 모형과 각 단일 분류기들을 배깅 방식으로 결합한 의사결정나무

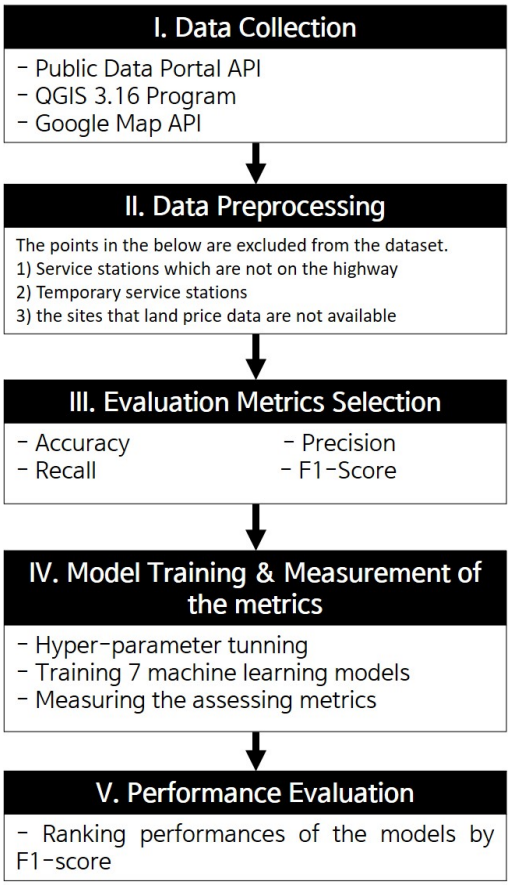

Figure 3. Research Design and Steps

배깅 앙상블(랜덤 포레스트), 서포트벡터머신 앙상블, K-최근접이웃법 배깅 앙상블 모형과 단일 분류기들을 모두 기저 분류기로 사용하는 보팅 앙상블 모형 총 7개의 머신러닝 모형을 활용한다. 그리고 이러한 모형들의 분석 결과에 대해 여러 성능 평가 척도를 이용해 모형들의 성과를 측정하고 이들 결과를 비교한다. 이러한 연구설계 및 절차를 그림으로 나타내면 다음의 Figure 3과 같다.

4. 분석 결과

4.1. 하이퍼파라미터 튜닝 및 분석절차

머신러닝 모형들의 학습 및 예측력 검증을 위해 본 연구에서는 각 모형이 최적의 성과를 낼 수 있는 파라미터를 탐색 및 설정하는 하이퍼 파라미터 튜닝을 실시하였다. 특히, 본 연구에서는 하이퍼 파라미터 튜닝의 방법으로 랜덤서치(Random Search)를 사용하였다. 랜덤서치는 무작위성(Randomness) 혹은 확률(Probability)에 기반하는 알고리즘으로 최적의 파라미터를 탐색할 때 무작위로 선정하는 특징을 가진다(Zabinsky 2009). 랜덤서치는 최적의 파라미터를 찾는 데 있어 높은 성능을 보여주는 것으로 나타났으며, 무작위적인 특성으로 인해 예외적인 파라미터 조합들을 생성함으로써 모형의 성과가 가장 좋은 최적의 파라미터를 효율적으로 찾을 수 있다는 장점이 있다(Bergstra and Bengio 2012). 하이퍼 파라미터 튜닝의 결과로 각 모형의 최적 성과를 산출할 수 있는 모형별 하이퍼 파라미터(Hyper-Parameter)는 Table 5에 제시되어 있다.

이러한 하이퍼 파라미터 튜닝 결과를 바탕으로 머신러닝 모형들의 파라미터를 설정한 상태에서 분석하였다. 이때, 배깅 앙상블 모형은 각각 기저 분류기들을 의사결정나무, 서포트벡터머신, k-최근접이웃법을 사용하여 구성하였으며, 보팅 앙상블 모형은 의사결정나무, 서포트벡터머신, K-최근접이웃법을 모두 포함해서 예측하였다.

또한, 전체 데이터를 학습용 데이터(Training Data)와 테스트용 데이터(Testing Data)를 가각 75%와 25%의 비율로 나누어 학습용 데이터를 통해 본 연구에서 사용된 모형들을 학습한 후 테스트용 데이터를 활용해 모형의 성능을 평가 및 비교하였다. 하이퍼 파라미터 튜닝과 머신러닝 모형의 구성 및 분석은 Python 프로그램의 Scikit-Learn 0.24.1 라이브러리

Table 5. Hyper-Parameters of Each Model

Model	Hyper-Parameter
Deicision Tree	criterion : gini, max_depth : 29, max_leaf_node : 65, min_sample : 6, min_samples_split : 0.02537318598749547
K-Nearest Neighborhood	weights : distance, n_neighbors : 5, metric : euclidean, leaf_size : 46, algorithm : auto
Support Vectcor Machine	C : 68.1397238572758, gamma : 1.2527258003249279e-05, kernel : rbf
Random Forest	criterion : gini, max_depth : 57, max_features : auto, min_samples_leaf : 27, min_samples_split : 0.1639771150222418, n_estinators : 25
KNN-Bagging Ensemble	n_estimators : 11, bootstrap_features : True, base_estimator__weights : distance, base_estimator__n_neights : 5, base_estimator__metric : manhattan
SVM-Bagging Ensemble	base_estimator__C : 39.8905781603212, base_estimator__gamma : 0.0009843278670203955, base_estimator__kernel : rbf, bootstrap : True bootstrap_features : True, max_samples : 0.7883067243096491, n_estimators : 1
Voting Ensemble	DT__max_depth : 65, DT__max_leaf_nodes : 74, DT__min_samples_leaf : 7, DT__min_samples _split : 0.11980654116261235, knn__metric : manhattan, knn__n_neighbors : 408, knn__wieghts : distance, svm__C : 8.874124299714731, svm__gamma : 0.27877197560853606, svm__kernel : sigmoid, voting : soft

Table 6. Confusion Matrix (Simple Classifier Models)

DT			SVM			KNN		
Predict / Actual	Positive	Negative	Predict / Actual	Positive	Negative	Predict / Actual	Positive	Negative
Positive	54	5	Positive	41	18	Positive	53	6
Negative	2	111	Negative	3	110	Negative	5	108

를 사용하였다.

4.2. 모형별 분석 결과 비교

하이퍼 파라미터 튜닝 결과를 바탕으로 모형들의 성능을 평가하기 위한 테스트용 데이터의 분석 결과를 제시하면 다음과 같다.

먼저 평가척도 측정의 기초가 되는 혼돈행렬을 모형별로 나타내면 다음의 Table 6~7과 같다. 혼돈행렬의 구성요소를 살펴보면 의사결정나무와 K-최근접이웃법의 경우에는 단일 머신러닝으로 분류하였을 때에 비해 실제값에 대하여 예측이 잘못된 것을 의미하는 FP와 FN의 값이 감소하였다. 이것은 앙상블 모형을 구성하였을 때 단일 분류기에 비해 그 예측 성과가 개선된 것임을 의미한다. 반면, 서포트벡터머신은 앙상블 모형을 구성하였을 때 FN의 값이 증가하여 단일 분류기에 비해 예측 성과가 악화되는 결과를 보여주었다. 보팅 앙상블의 경우에는 의사결정나무로만 예측을 실시하였을 경우를 제외하고는 모든 단일 머신러닝 모형에 비해 분류 성과가 좋았으며, 랜덤포레스트 다음으로 분류 성능이 뛰어난 것으로 나타났다.

Table 7. Confusion Matrix (Ensemble Classifier Models)

Random ForestRandom Forest			SVM-Bagging		
Actual \ Predict	Positive	Negative	Actual \ Predict	Positive	Negative
Positive	54	5	Positive	26	33
Negative	1	112	Negative	0	113
KNN-Bagging			**Voting**		
Actual \ Predict	Positive	Negative	Actual \ Predict	Positive	Negative
Positive	53	6	Positive	54	5
Negative	1	112	Negative	5	108

Table 8. Accuracy, Precision, Recall, F1-Score

Model \ Metrics	Accuracy	Precision	Recall	F1-Score
DT	0.9593	0.9643	0.9153	0.9391
SVM	0.8779	0.9318	0.6949	0.7961
KNN	0.9360	0.9138	0.8983	0.9060
Random Forest	0.9652	0.9818	0.9153	0.9474
SVM-Bagging	0.8256	1.0000	0.4915	0.6591
KNN-Bagging	0.9593	0.9814	0.8983	0.9381
Voting	0.9419	0.9153	0.9153	0.9153

정확도, 정밀도, 재현율, F1-점수는 Table 7에 제시되어 있다. 단일 머신러닝 모형을 활용해 분석을 하였을 때와 비교해 앙상블 모형을 구성했을 경우 성과의 변화 모습이 달랐다. 의사결정나무는 앙상블 모형으로 구성한 렌덤포레스트의 성과가 재현율을 제외한 모든 성과 측정치에서 개선되었다. 또한, K-최근접이웃법의 경우에도 앙상블 모형을 구성하였을 때 전반적인 성과가 개선되었다. 서포트벡터머신의 경우에는 앙상블 모형을 구성하여 분석하였을 경우 오히려 성과가 악화되었다. 보팅 앙상블 모형의 경우 의사결정나무 모형을 제외한 나머지 단일 머신러닝 모형에 비해 전반적으로 성능이 개선되었다.

전술한 바와 같이 모형의 성과를 F1-점수에 따라 평가하면 랜덤포레스트가 가장 성능이 우수한 것으로 나타났으며, 의사결정나무, 보팅 앙상블 모형의 순으로 성능이 우수하였다. 이러한 결과는 랜덤포레스트가 부트스트랩(Bootstrap) 기법을 사용하여 무작위로 표본을 생성하고 독립변수에 대한 무선표집(Random Sampling)을 실시함으로써 기저 분류기인 의사결정나무 간의 독립성을 확보해 예측오차를 줄이는 특성에서 기인하는 것으로 판단된다(유진은 2015). 이는 머신러닝 모형을 통한 예측을 실시할 때 머신러닝 모형과 더불어 모형의 학습 및 테스트에 활용하는 데이터에 무작위성을 부여하는 방법 등의 방법에 대한 고

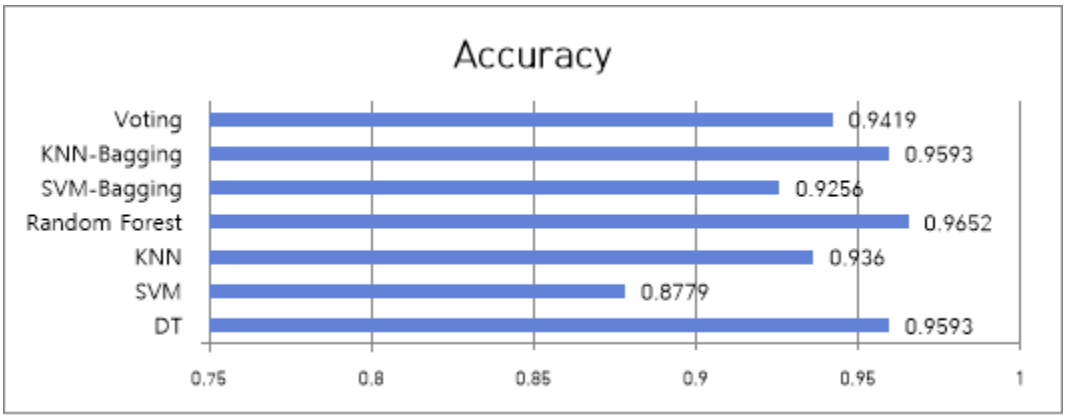

Figure 4. Accuracy Scores

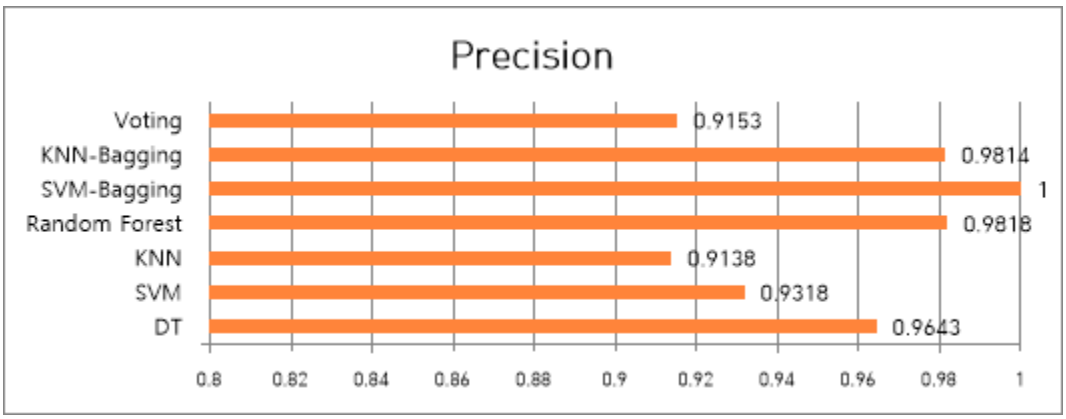

Figure 5. Precision Scores

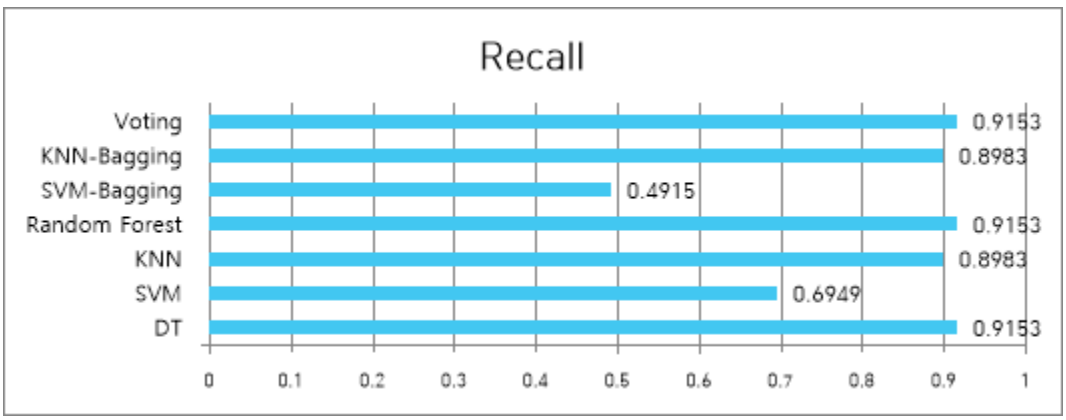

Figure 6. Recall Scores

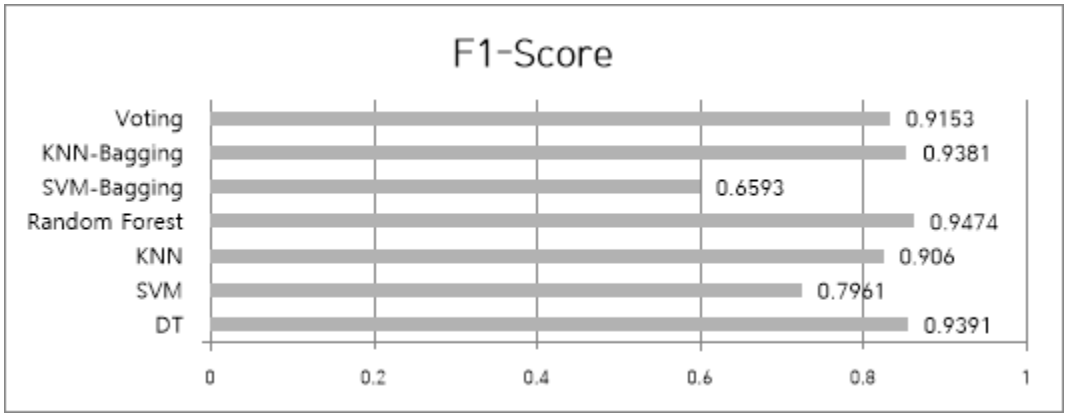

Figure 7. F1-Scores

찰도 필요함을 시사한다. 이러한 예측 결과들을 그래프로 나타내면 다음의 Figures 4~7과 같다.

5. 결론

5.1. 연구요약과 시사점

미세먼지 등으로 인한 대기오염, 온실가스 배출에 따른 지구온난화 등 환경오염 문제가 대두됨에 따라, 기존의 화석연료를 대체할 새로운 청정에너지원의 필요성이 증대되고 있다. 이와 같은 수요에 따라 세계 각국에서는 청정에너지인 수소를 기반으로 하는 수소경제로의 전환을 위한 노력이 이루어지고 있으며, 국내에서도 정부 차원에서 수소 관련 분야를 선도하고자 정책적으로 많은 투자를 하고 있다.

이러한 맥락에서 물류 분야에서도 수소를 도입하고자 수소추진화물차의 개발과 수소충전소 설치 등 수소 기반 물류 네트워크 구축을 위한 노력이 수행되고 있다. 본 연구에서는 이 중 수소충전소 최적 입지 문제를 다루고자 하였는데, 수소 충전 인프라는 수소 기반의 물류 체계를 구축함에 있어 핵심적인 요소이고 성공적인 수소충전소의 입지 결정은 수소자동차의 보급을 촉진하는데 기여할 수 있다는 사실에 기반하고 있다.

기존에 최적화 기법만을 사용하여 수소충전소 최적입지를 결정하는 선행연구들과는 달리, 본 연구에서는 최적 입지를 결정하는데 있어 최적화 기법의 중요한 투입변수 중 하나인 충전소 후보지에 대해 머신러닝 모형들을 통해 예비적 차원의 검토를 수행하고 그 활용가능성을 확인하였다. 머신러닝 분야에서 수소충전소의 최적 입지를 선정하는 문제는 적절한 충전소 후보지 여부를 결정하는 것이므로 분류(Classification) 문제에 해당한다. 본 연구에서는 선행연구에서 분류 문제에 탁월한 성능을 보이는 것이 증명된 머신러닝 모형들을 분석에 사용하였다. 모형들을 학습하기 위한 데이터의 변수로 개별공시지가, 수소공급지와의 거리, 영업소와의 거리, 지점별로 최단 지점에 위치한 5곳과의 최단거리를 독립변수로 하여 분석을 실시하였다.

분석 결과, 단일 머신러닝 모형으로 수소충전소의 최적 입지를 예측하였을 때보다 앙상블 모형을 구성한 경우 그 성능의 변화 양상에 차이가 존재했다. 의사결정나무와 K-최근접이웃법의 경우 앙상블 모형에서 성과가 전반적으로 개선되었지만, 서포트벡터머신은 앙상블 모형을 구성하였을 때 성과가 악화된 것으로 나타났다. 머신러닝 모형들의 성과를 보수적인 관점에서 평가하고자 F1-점수를 기준으로 모형들의 성능

을 평가하였을 때 랜덤포레스트 모형이 가장 우수하였으며, 의사결정나무, 보팅 모형 순으로 성능이 우수하였다.

본 연구는 다음과 같은 학문적·실무적 시사점을 가진다. 학문적인 관점에서 머신러닝 모형들을 사용하여 수소충전소의 최적 입지를 결정을 위한 충전소 후보지에 대한 예비적 검토를 함으로써 추후 최적화 기법을 적용 시 그 중요 투입변수인 충전소 후보지에 다양한 공간적 특성을 고려하였다. 최적 입지를 선정하는 문제는 전통적으로 최적화 기법만을 활용하여 연구가 수행되어 왔으며, 수소충전소의 위치를 결정하는 것을 주제로 한 선행연구들도 대부분 최적화 기법만을 중심으로 연구를 수행하였다. 이러한 연구 동향과는 달리, 본 연구는 다양한 분야에서 그 유용성이 입증된 머신러닝 모형을 수소충전소의 최적 입지 문제에 적용하여 충전소 후보지에 대한 공간적인 검토가 이루어졌다는 점에서 의의가 있다.

실무적인 관점에서는 본 연구가 실제 정부의 수소 관련 정책에 따라 연구를 설계 및 수행하였다는 것이다. 현재 정부는 고속도로의 수소충전소 후보지로 고속도로 휴게소들을 그 대상으로 하고 있으며, 이에 따라 지역 간의 수소자동차 통행을 뒷받침하여 수소자동차 보급을 촉진하는 것을 목표로 하고 있다. 본 연구에서는 이러한 정부 정책에 맞추어 고속도로 휴게소를 잠재적인 후보지로 데이터를 구성 및 분석하여 실무적인 유용성이 있다는 점에서 그 의의가 있다. 또한 본 연구는 머신러닝 모형을 활용하여 특정 지점과 관련된 다양한 변수들을 학습 및 테스트하여 최적 입지 결정을 위한 예비적인 검토를 수행하는 연구라는 점에서, 최근 이슈가 되고 있는 COVID-19 선별 진료소와 같은 시설들의 최적 입지를 결정하는데도 유용하게 활용될 수 있다는 실무적 이점을 가지고 있다.

5.2. 한계점과 향후 연구 방향

이러한 시사점들에도 불구하고 본 연구는 다음과 같은 한계점들을 가진다. 먼저 화물차 전용 수소충전소의 최적입지를 결정하고자 하였으나, 일반 자동차에 기반한 데이터들을 바탕으로 모형의 학습 및 예측을 실시하였다. 이는 수소충전소와 관련된 분야가 비교적 최근에 연구가 이루어지기 시작하였으며, 현재 국내 수소 산업의 구축이 초기 단계에 머물러 있어 선행연구와 관련 데이터의 수집이 어려운 점에 기인하였다. 그러므로 후속 연구에서는 화물차나 수소추진 화물차와 더욱 관련성이 높은 데이터를 구성하여 연구를 수행해야 할 것이다.

다음으로 본 연구는 수소충전소 최적 입지를 선정함에 있어 모형 간의 성능을 단순 비교함으로써 머신러닝 모형들의 해당 연구 주제에 적용가능성만을 확인하였다는 것이다. 본 연구는 정부 정책에 따라 고속도로 휴게소만을 충전소 설치 후보지로만 고려하였기에, 모형의 학습에 필요한 데이터뿐 아니라 실제 학습된 모형을 토대로 예측을 실시할 부지 데이터의 수집에 어려움이 있었다. 그러므로 후속 연구에서는 고속도로상의 유휴부지 등을 실제 예측에 활용할 데이터로 하여 실제로 충전소 후보지를 결정할 수 있도록 할 필요가 있다.

또한, 본 연구는 수소충전소의 최적 입지를 결정하는데 있어 고속도로상 지점들의 위치 및 거리에 기반한 변수와 수소충전소 인프라와 관련된 변수만을 사용하였고, 통행 패턴, 수소 충전 특성 등 수소추진 자동차 사용자 관점의 요소들을 고려하지 못하였다. 그러므로 후속 연구에서는 이러한 사용자 입장에서의 요소들을 고려한 연구를 수행하는 것이 바람직할 것이다.

마지막으로 본 연구에서 고속도로 휴게소 및 무작위 지점과 관련된 독립변수의 수가 제한적인 한계점이 있다. 본 연구는 데이터에 포함된 지점들의 위치와

관련된 변수만을 중심으로 예측변수군을 구성하여 수소충전소의 종류 등 수소 공급 네트워크의 다양한 측면들과 관련된 변수들을 고려하지 못하였다. 따라서 후속 연구에는 수소 공급 네트워크를 종합적인 관점에서 평가할 수 있도록 추가 변수들을 예측변수로 활용한다면 더욱 의미 있는 연구가 될 것으로 기대된다.

참고문헌
References

국토교통부. 2019. 수소 인프라 및 충전소 구축방안 [인터넷].
[http://www.molit.go.kr/USR/NEWS/m_71/dtl.jsp?lcmspage=1&id=95082944]]. Last accessed 20 September 2021.

김거중, 박준식, 고승렬. 2019. 하이브리드형 수소공급 방식을 고려한 수소충전소 입지 선정 모형 개발. 교통연구. 26(2):53-70.

Kim GJ, Park JS, Go SR. 2019. Location problem of hydrogen refueling station considering hydrid hydrogen supply system. *Journal of Transport Research*. 26(2):53-70.

김남진, 배영철. 2018. K-최근접 이웃 알고리즘을 적용한 펌프와 모터의 상태 진단. 한국전자통신학회 논문지. 13(6):1249-1256.

Kim NJ, Bae YC. 2018. Status Diagnosis of Pump and Motor Applying K-Nearest Neighbors. *The Journal of The Korea Institute of Electronic Communication Sciences*. 13(6):1249-1256.

김봉진, 국지훈, 조상민. 2014. 지리정보시스템을 이용한 고속국도에서의 수소충전소 구축 방:안. 한국수소 및 신에너지학회 논문집. 25(3): 255-263.

김수환, 류준형. 2020. 수소충전소 최적 위치 선정을 위한 기계 학습 기반 방법론. Korean Chmical Engineering Research. 58(4):573-580.

Kim SH, Ryu JH. 2020. A Machine Learning based Methodology for Selecting Optimal Location of Hydrogen Refueling Stations. *Korean Chmical Engineering Research*. 58(4):573-580.

김은경, 전명식, 방성완. 2016. 그룹변수를 포함하는 불균형 자료의 분류분석을 위한 서포트 벡터 머신. 응용통계연구. 29(5):961-975.

Kim EK, Jhun MS, Bang SW. 2016. Hierarchically penalized support vector machine for the classi cation of imbalanced data with grouped variables. *The Korean Journal of Applied Statistics*. 29(5):961-975.

김은미, 김상봉, 조은서. 2020. 기계학습을 활용한 주택매도 결정요인 분석 및 예측모델 구축. 지적과 국토정보. 50(1):181-200.

Kim EM, Kim SB, Cho ES. 2020. Using Mechanical Learning Analysis of Determinants of Housing Sales and Establishment of Forecasting Models. *Journal of Cadastre & Land InformatiX*. 50(1): 181-200.

김정화. 2020. 수소경제 활성화 로드맵 달성에 따른 교통 부문의 대기오염원 저감 효과 분석.한국수소 및 신에너지학회논문집. 31(6): 522-529.

Kim JH. 2020. Air Pollutant Reduction Effect on Road Mobility in Hydrogen Economy Era. *Transactions of the Korean Hydrogen and New Energy Society*. 1(6): 522-529.

김한민. 2020. 앙상블 머신러닝 기법과 블록체인 정보를 활용한 이더리움 엉클 블록 예측 분석. 디지털융복합연구. 18(11):129-136.

Kim HM. 2020. Predictive Analysis of Ethereum Uncle Block using Ensemble Machine Learning Technique and Blockchain Information. *Journal of Digital Convergence*. 18(11):129-136.

대한민국 정책브리핑. 2020. 수소경제 [인터넷].

[https://www.korea.kr/special/policyCurationView.do?newsId=148857966] Last accessed 22 September 2021.

민성환. 2016. 부도예측을 위한 KNN 앙상블 모형의 동시최적화. 지능정보연구. 22(1): 139-157.

Min SH. 2015. Investigating Dynamic Mutation Process of Issues Using Unstructured Text Analysis. *Journal of Intelligence and Information Systems*. 22(1): 139-157.

박종훈, 최병인, 이종훈. 2006. 밀도에 기반한 퍼지 서포트 벡터 머신을 이용한 멀티 카데고리에서의 패턴 분류. 한국지능시스템학회 학술발표 논문집. 16(2):251-254.

Park JH. Choi BI, Rhee CH. 2006. Density based Fuzzy Support Vector Machines for multi-category Pattern Classification. *Proceedings of KFIS Autumn Conference*. 16(2):251-254.

박지만, 조두영, 이상선, 이민섭, 남한식, 양혜림. 2018. 인공지능과 국토정보를 활용한 노인복지 취약지구 추출방법에 관한 연구. 지적과 국토정보. 48(1):169-186.

Park JM, Cho DY, Lee SS, Lee MS, Nam HS, Yang HR. 2018. A Study on The Methodology of Extracting the vulnerable districs of the Aged Welfare Using Artificial Intelligence and Geospatial Information. *Jounral of Cadastre & Land InformatiX*. 48(1):169-186.

외교부. 2015. 기후변화협상. [https://www.mofa.go.kr/www/wpge/m_20150/contents.do]. Last accessed 1 September 2021.

유진은. 2015. 랜덤 포레스트. 교육평가연구. 28(2): 427-448.

Yoo JE. 2015. Random forest, an alternative data mining technique to decision tree. *Journal of Educational Evaluation*. 28(2):427-448.

이용준, 선종완. 2020. XGBoost를 활용한 고속도로 콘크리트 포장 파손 예측. 한국건설관리학회 논문집. 21(6):46-55.

Lee YJ, Sung JW. 2020. Predicting Highway Concrete Pavement Damage using XGBoost. *Korean journal of construction engineering and management*. 21(6):46-55.

이우진, 전준헌. 2021. 머신러닝을 활용한 공공자전거 대여소의 입지 예측에 대한 연구. 한국IT정책경영학회 논문지. 13(4):2553-2559.

Lee WJ, Jeon JH. 2021. A Study on Prediction of the Location of Public Bicycle Rental Stations Using Machine Learning. *Journal of The Korea Society of Information Technology Policy & Management*. 13(4):2553-2559.

이창한, 성춘자. 2018. 공시지가산정을 위한 지형 · 지세조사 자료의 정확도 분석. 지적과 국토정보. 48(1):153-167.

Lee CH, Sung CJ. 2018. Accuracy Analysis of Topographic Survey Data for the Official Land Price Appraising. *Journal of Cadstre & Land InformatiX*. 48(1):153-167.

전병욱, 강지수, 정경용. 2021. 도로교통 이머징 리스크 탐지를 위한 AutoML과 CNN 기반 소프트 보팅 앙상블 분류 모델. 융합정보논문지. 11(7): 14-20.

Jeon BU, Kang JS, Chung KY. 2021. AutoML and CNN-based Soft-voting Ensemble Classification Model For Road Traffic Emerging Risk Detection. *Journal of Convergence for Information Technology*. 11(7):14-20.

정기대. 2019. 수소경제의 경제적·기술적 이슈 – Value Chain 5단계 중심 -. 10(5):1-12.

정용찬, 류혜연, 이수정, 서동주, 박찬건. 2021. 머신러닝 기반의 KSORAS 재범요인 확인 연구 : 의사결

정나무 분석과 랜덤포레스트 기법을 활용하여. 한국경찰연구. 20(1):323-350.

Jeng YC, Ryu HY, Lee SJ, Seo DJ, Park CG. 2021. Identification recidivism risk factors study based on machine learning: Using decision tree analysis and random forest algorithm. *Korean Police Studies Review*. 20(1):323-350.

Alazemi J, Andrews J. 2015. Autumotive hydrogen fuelling station: An international review. *Renewable and Sustainable Energy Reviews*, 48:483-499.

Bergstra J, Bengio Y. 2012. Random search for hyper-parameter optimization. *Journal of machine learning research*. 13(2):281-305.

Brandon N P, Kurban Z. 2017. Clean energy and the hydrogen economy. Philosophical Transactions of the Royal Society A: Mathematical. *Physical and Engineering Sciences*. 375(2098): 1-17.

Bühlmann P. 2012. *Bagging, boosting and ensemble methods. In Handbook of computational statistics*. USA: Springer, 985-1022,

Cervantes J, Garcia-Lamont F, Rodríguez-Mazahua L, and Lopez A. 2020. A comprehensive survey on support vector machine classification: Applications, challenges and trends. *Neurocomputing*, 408:189-215.

Dagdougui H. 2012. Models, methods and approaches for the planning and design of the future hydrogen supply chain. *International Journal of hydrogen Energy*. 37(6):5318-5327.

Damavandi H, Abdolvand N, Karimipour F. 2019. Utilizing location-based social network data for optimal retail store placement. *Earth Obsevation and Geomatics Engineering*. 3(2): 77-91.

Elloumi S, Labbé M, Pochet Y. 2004. A new formulation and resolution method for the p-center problem. *INFORMS Journal on Computing*. 16(1):84-94.

Frade I, Ribeiro A, Gonçalves G, Antunes A P. 2011 Optimal location of charging stations for electric vehicles in a neighborhood in Lisbon, Portugal. *Transportation Research Record*. 2252(1):91-98.

Friedl M A, Brodley C E 1997. Decision tree classification of land cover from remotely sensed data. *Remote Sensing of Environment*. 61(3):399-409.

Gaikwad D P, Thool R C. 2015. Intrusion detection system using bagging ensemble method of machine learning. *In 2015 international conference on computing commnication control and automation*, 291-295.

Gandhi I, Pandey M. 2015. Hybrid ensemble of classifiers using voting. *In 2015 international conference on green computing and Interne of Things (ICGCIoT)*, 399-404.

Gazalba I, Reza N G I. 2017. Comparative analysis of k-nearest neighbor and modified k-nearest neighbor algorithm for data classification. *In 2017 2nd International conferences on Information Technology, Information Systems and Electrical Engineering (ICITISEE)*. 294-298.

Gielen D, Gorini R, Wagner N, Leme R, Gutierrez L, Prakash G, Renner M. 2019. Global energy transformation: a roadmap to 2050. *International Renewable Energy Agency (IRENA)*. 1-52.

Hu L Y, Huang M W, Ke S W, Tsai C F, 2016. The

distance Function effect on k-nearest neighbor classification for medical datasets, SpringerPlus. 5(1):1-9.

Ozbas E E, Aksu D, Ongen A, Aydin M A, and Ozcan H K. 2019. Hydrogen production via biomass gasification, and modeling by supervised machine learning algorithms. *International Journal of Hydrogen Energy*. 44(32): 17260-17268.

Hosseini M, MirHassni S A. 2015. Refueling-station location problem under uncertainty. *Transportation Research Part E: Logistics and Transportation Review*. 84:101-116.

Jeong I J. 2017. An optimal approach for a set covering version of the refueling-station location problem and its application to a diffusion model. *International Journal of Sustainable Transportation*. 11(2):86-97.

Kaur H, Pannu H S, and Malihi A K. 2019. A systematic review on imbalanced data challenges in machine learning: Applications and solutions. *ACM Computing Surveys (CSUR)*. 52(4):1-36

Kim H C, Pang S, Je H M, Kim D, and Bang S Y. 2002. *Support vector machine ensemble with bagging*. Berlin, Heidelberg: Springer, p. 397-408.

Kim J H, Ki B S, Savarese S. 2012. Comparing image classification methods: K-nearest-neighbor and support-vector-machines. *In Procddings of the 6th WSEAS international conference on Computer ENginering and Applications, and Proceedings of the 2012 American conference on Applied Mathematics*. 1001:48109-2012.

Kluschke P, nann T N, Plötz P, Wietschel M. 2019. Market diffusion of alternative fuels and powertrains in heavy-duty vehicles: a literature review. *Energy Report*. 5: 1010-1024.

Ko J, Gim T H T, Guensler R. 2017. Locating refuelling stations for alternative fuel vehicles: a review on models and applications. *Transport Reviews*. 37(5): 551-570.

Kuby M, Lines L, SchultzR, Xie Z, Kim J G, Lim S. 2009. Optimization of hydrogen stations in Florida using the flow-refueling location model. *International journal of hydrogen energy*. 34(15):6045-6064.

Li L, Manier H, Manier M A. 2019. Hydrogen supply chain network design: An optimization-oriented review. *Renewable and Sustainable Energy Reviews*. 103:342-360.

Lin R H, Ye Z Z, Wu B D. 2020. A review of hydrogen station location models. *International Journal of Hydrogen Energy*. 45(39):20176-20183.

Lin Z, Ou S, Elgowainy A, Reddi K, Veenstra M, Verduzco L. 2018. A Method for determining the optimal delivered hydrogen pressure for fuel cell electric vehicles. *Applied Energy*. 261(15):183-194.

Muratori M, Bush B, Hunter C, Melaina M W. 2018. Modeling hydrogen refueling infrastructure to support passenger vehicles. *Energies*. 11(5): 1-14.

Pradhan A. 2012. Support vector machine-a survey. *International Journal of Emerging Technology and Advanced Engineering*. 2(8):82-85.

Priyam A, Abhijeeta G R, Rathee A, Srivastava S. 2013. Comparative analysis of decision tree

classification algorithms. *International Journal of current engineering and technology*. 3(2):334-337.

Shah S A A, Aziz W, Arif M, Nadeem M S A. 2015. Decision trees based classification of cardiotocograms using bagging approach. *In 2015 13th international conference on frontiers of information technology(FIT)*. 12-17.

Sharma H, Kumar S. 2016. A survey on decision tree algorithms of classification in data mining. *International Journal of Science and Research (IJSR)*. 5(4):2094-2097.

Sharma D, Kumar N. 2017. A review on machine learning algorithms, tasks and applications. *International Journal of Advanced Research in Computer Engineering & Technology (IJARCET)*. 6(10):1548-1552.

Teng X, Gong Y. 2018. Research on application of machine learning in data mining. *In IOP conference series: materials science and engineering*. 392(6):1-5.

Wang H, Ma C, Zhou L. 2009. A brief review of machine learning and its application. *In 2009 international conference on information engineering and computer science*, 1-4.

Zabinsky Z B. 2009. *Random search algorithms*. USA: Department of Industrial and Systems Engineering, University of Washington. p. 1-16.

Zaheer N, Hassan S U, Ali M, Shabbir M. 2021. Optimal school site selection in Urban areas using deep neural networks. *Journal of Ambient Intelligence and Humanized Computing*. 1-15.

2021년 10월 05일 원고접수(Received)
2021년 10월 28일 1차심사(1st Reviewed)
2021년 11월 25일 게재확정(Accepted)

초 록

대기오염, 지구온난화 문제 등 환경 문제의 심각성이 대두되면서 청정 연료의 관심이 커지고 있다. 그 중 수소는 기존 화석연료와는 달리 연소 시 부산물로 수분만이 발생하는 대표적인 친환경 에너지원으로 현재 다양한 분야에서 주목을 받고 있다. 물류 분야에서도 수소를 활용한 물류 네트워크를 구축하기 위해 다양한 정책적 노력이 활발히 이루어지고 있다. 이러한 수소 물류 네트워크의 구축에 있어 수소충전소의 입지 결정은 매우 중요한 문제이다. 최근 개발된 수소추진(수소연료전지) 화물차에 수소를 공급하는 충전소는 수소 기반 물류체계가 본격적으로 자리 잡는 데 있어 필수 불가결한 요소이다. 이러한 수소충전소의 최적 입지를 결정하는 선행연구는 대부분 수리적 모형에 기반한 최적화 기법만을 사용하여 수소충전소의 최적 입지를 결정하고자 하였다. 본 연구에서는 기존 연구의 동향과는 차별화하여 최적화 기법의 중요한 투입 변수 중 하나인 충전소 후보지에 대한 공간적 특성을 검토하는 방법으로 머신러닝 모형들을 활용하고 그 적용가능성을 확인하였다. 머신러닝은 다양한 분야에서 우수한 성과를 증명한 기법이지만 수소충전소의 최적 입지를 결정하는 연구 분야에서는 아직 적용된 바가 없다. 이를 위해 본 연구에서는 개별공시지가, 수소공급지와의 거리 등 전국 고속도로 휴게소와 고속도로의 무작위 지점들의 위치와 관련된 변수들을 독립변수로 선정하여 단일 머신러닝 모형과 앙상블 모형을 적용하고 그 성과를 비교하였다. 분석 결과, 랜덤포레스트(Random Forest) 모형이 가장 우수한 성과를 보였으며, 다른 모형들 또한 우수한 분류 성능을 보여 최적 입지 문제에 대해 공간적 특성을 예비적으로 검토하는 방법론으로써 머신러닝의 적용 가능성을 확인할 수 있었다. 따라서 머신러닝 모형은 수소충전소의 최적 입지 결정 분야에서 향후 최적화 기법을 적용한 연구의 예비적 검토 방법론으로 널리 활용할 수 있을 것으로 기대된다.

주요어 : 머신러닝, 수소 물류 네트워크, 수소충전소, 최적입지 결정, 예비적 검토 방법론