XAI 기반 2025년도 KTX 수요 예측 및 정책적 의사결정

차명주a, 오영택b, 이승연b, 김경원a,*∗*

*a인천대학교 글로벌정경대학 무역학부*

*b한국철도공사 철도연구원*

# Abstract

고속철도는 전기차보다 12배, 내연기관 차량보다 26배 낮은 CO2 배출량을 기록하며, 지속 가능한 교통수단의 대표 사례로 자리 잡고 있다. 고속철도의 수요 변화를 정밀하게 예측하는 것은 단순한 운영 효율성 개선을 넘어, 장기적인 교통 정책 수립과 지속 가능한 인프라 구축을 위한 핵심 요소로 작용한다. 본 연구는 고성능 인공지능 알고리즘과 설명 가능한 인공지능(XAI, Explainable AI)을 활용하여 2025년도 KTX 수요를 정밀하게 예측하고, 예측 결과를 바탕으로 실질적인 비즈니스 및 정책적 의사결정을 지원하는 것을 주된 목적으로 한다. 대표적인 AI 알고리즘을 활용하여 KTX 수요 예측의 오류를 최대 2.49%까지 낮추며 정확도를 대폭 향상시켰다. 또한, 대표적 XAI 기술인 SHAP 알고리즘을 활용하여 예측 결과와 변수들의 기여 정도를 시각화 함으로써, 비즈니스적 정책 설계 및 자원 배분 의사결정 과정에 신뢰도를 높였다. KTX 노선에 따라 기여하는 변수들은 다양하게 변화될 수 있는 가능성을 제시하며 실시간 활용가능한 비즈니스 애널리틱스의 방법론의 필요성을 확인하였다. 최종적으로 2025년도 KTX 수요는 작년보다는 최대 9.45% 정도 감소할 수 있지만, 코로나 이전보다 최대 18.47%까지 상승하는 수치로 수요가 점진적으로 증가되며 안정화 될 것으로 예상된다.

*Keywords:* Demand Forecasting of KTX, Explainable Prediction, Machine and Deep Learning, SHAP, Decision Support

*∗*Corresponding author: thekimk.kr@gmail.com

*ORCID(s):* 0000-0001-6530-8426 (K. Kim)

# Introduction

고속철도는 한국을 포함한 전 세계 교통 시스템에서 지속 가능한 발전의 핵심 축으로 자리 잡고 있다. 한국의 KTX, 프랑스의 TGV, 일본의 신칸센, 중국의 HSR과 같은 사례는 고속철도가 국가 경제와 사회적 연결성 강화에 미치는 긍정적 영향을 잘 보여준다. 국외 사례를 살펴보면, 특히 중국에서는 고속철도가 도시 간 연결성을 증대시키며 고용 증가, 고정 자산 투자 확대, 평균 임금 상승 등 지역 경제 성장에 긍정적인 영향을 미치는 것으로 나타났다 (Lin, Dhakal, and Wu, 2021). 또한, 국내에서도 고속철도의 개통은 지역 간 이동 시간을 단축하고, 인구 및 사업체 증가에 긍정적인 영향을 미치며, 국가 간 경제적 연결성을 확대하는 주요 인프라로 기능하는 것으로 나타났다 (유현아 외, 2024). 이들 사례는 고속철도가 단순히 국가 내 이동 수단에 국한되지 않고, 경제적 연결성과 국제적 경쟁력을 강화하는 주요 인프라로 기능할 수 있음을 입증한다.

현재 25개국 이상이 고속철도를 도입하여 주요 교통수단으로 활용하고 있으며, 이러한 국가는 고속철도를 통해 물류와 인구의 이동성을 극대화하며 교통체계의 효율성과 환경적 지속 가능성을 동시에 달성하고 있다. 특히, 고속철도는 전기차보다 12배, 내연기관 차량보다 26배 낮은 CO2 배출량을 기록하며, 지속 가능한 교통수단의 대표 사례로 자리 잡고 있다 (Union, 2023).

글로벌 고속철도의 발전 속에서 한국의 KTX는 대표적인 고속철도 시스템으로 자리 잡았다. 2022년 기준 약 7,500만 명의 이용객을 기록한 이후 지속적인 증가세를 보였으며, 2024년에는 연간 이용객 수가 총 1억 1,658만 명으로 전년 대비 5.4% 증가하며, KTX가 단순한 교통수단을 넘어 국가 기반 교통망의 중심축으로 기능하고 있음을 입증하고 있다 (국토교통부, 2024).

그러나, 이러한 이용객 증가에 따른 운영상의 도전 과제도 점차 복잡해지고 있으며 효과적으로 해결하기 위해서는 수요 변화를 사전에 예측하고, 이를 기반으로 최적화된 운영 전략을 수립하는 것이 필수적이다. 이에 따라, 국토교통부는 미래 교통 시스템의 효율성을 제고하기 위해 수요 예측과 연계된 미래 선도 기술 개발을 적극적으로 추진하고 있다 (기술과가치, 2020). 이러한 맥락에서, 고속철도의 수요 변화를 정밀하게 예측하는 것은 단순한 운영 효율성 개선을 넘어, 장기적인 교통 정책 수립과 지속 가능한 인프라 구축을 위한 핵심 요소로 작용한다. 그러나 효과적인 수요 예측을 위해서는 고속철도 이용 패턴을 정확히 반영할 수 있는 정교한 분석 기법이 필수적이며, 이에 대한 연구는 지속적으로 이루어져 왔다.

고속철도 수요 예측은 주로 전통적인 통계적 접근에 기반한 시계열 분석 기법을 활용해 이루어져 왔다. (Kim과 Kim, 2011)는 개입 ARIMA 모형을 사용하여 경부고속철도 2단계 개통과 같은 정책적 개입이 수요에 미치는 영향을 분석하였다. 이 연구는 정책 변화에 따른 단기적 수요 변화를 효과적으로 포착했지만, 열차 운임, 운행시간, 서비스 품질 등 다양한 변수 간의 상호작용을 반영하지 못했고, 데이터의 선형적 구조를 가정하여 설명력에 한계를 드러냈다. 이러한 한계는 고속철도 수요가 시간적, 공간적으로 복잡한 비선형적 특성을 가지며, 다양한 외부 요인의 영향을 받는다는 점에서 더욱 두드러진다. (Cha et al., 2019) 은 다중 개입 계절형 ARIMA 모형을 활용하여 고속철도 개통이나 국가 전염병과 같은 외부 환경변화를 반영한 수요 예측을 수행하였다. 계절적 요인과 외부 충격을 함께 고려해 이전보다 개선된 결과를 도출했지만, 전통적인 시계열 분석 기법이 가진 구조적 한계에서 벗어나지 못했다. 이러한 방법론은 데이터가 선형적 관계를 따르며, 정상성(stationarity)을 만족해야 한다는 전제 조건을 가지므로 현실 세계의 복잡한 데이터 특성을 충분히 반영하지 못하고 정보 손실 가능성을 내포하고 있다. 결과적으로, ARIMA와 같은 시계열 분석 기법은 과거 데이터를 기반으로 수요 예측에 널리 활용되어 왔으나, 선형적 관계를 기반으로 한다는 점에서 현실 세계의 복잡하고 비선형적인 수요 변화를 효과적으로 예측하기에는 한계가 있다(Sima et al., 2018). 이러한 전통적 통계 기반 기법은 대규모 데이터 처리와 실시간 분석이 요구되는 현대 고속철도 수요 예측 문제에서 특히 두드러진 한계를 보인다.

머신러닝 및 딥러닝과 같은 고성능 인공지능 알고리즘은 데이터 내에 내재된 복잡한 패턴을 학습하고, 변수 간의 관계를 자동으로 탐지하며, 대규모 데이터를 처리하는 데 강점을 지닌다. 특히, 인공지능 기반 수요 예측 기법은 높은 예측 정확도와 계산 효율성을 제공하며, 비선형 데이터 패턴의 학습과 다양한 변수 간의 관계 탐지가 가능하다는 점에서 기존 통계적 접근법의 한계를 효과적으로 보완한다 (Jeong과 Lim, 2019). 또한 최근에는 머신러닝 및 딥러닝을 활용하여 고속철도 수요 예측에서 더 높은 정확도를 달성하고, 효율적인 운영 전략 수립에 기여할 가능성을 제시해왔다. 예를 들어, LSTM(Long Short-Term Memory)과 XGBoost를 적용한 연구에서는 비선형 데이터 패턴을 효과적으로 학습하고, 기존 통계 기반 모델보다 더 높은 예측 성능을 보임으로써 알고리즘의 실효성을 입증하였다 (심진호 et al., 2024). 하지만 인공지능 알고리즘의 발전에도 불구하고 고속철도 수요예측에 활용되는 사례나 연구는 아직 부족한 실정이다.

인공지능의 알고리즘의 높은 성능에도 불구하고 구조가 매우 복잡하여 왜 그러한 결과가 도출되었는지 설명하지 못하는 한계가 있다. 이를 “블랙박스” 이슈라고도 하며 의사결정 과정에서 예측 결과의 신뢰성을 낮추고, 정책 설계나 자원 배분과 같은 실제 활용에 제약을 초래할 수 있다. 이러한 한계는 특히, 교통망 운영과 같은 대규모 인프라에서 심각한 문제로 작용할 수 있다. 이와 같은 문제를 해결하기 위해 최근에는 설명 가능한 인공지능(XAI, Explainable AI)이 주목받고 있다. 설명 가능한 인공지능(XAI)는 인공지능 시스템이 수행하는 예측 및 의사결정 과정을 인간이 이해할 수 있도록 설명하는 기술로, AI 시스템의 행동과 상태를 명확히 전달하여 신뢰성을 높이는 것을 목표로 한다 (David et al., 2019). 이 기술은 금융, 의료, 제조 등 다양한 분야에서 활용 사례를 통해 그 중요성을 입증하고 있다. 예를 들어, 금융 분야에서는 SHAP(Shapley Additive Explanations)와 LIME(Local Interpretable Model-agnostic Explanations) 같은 도구를 활용하여 신용등급 평가와 대출 의사결정에서 예측 결과의 해석 가능성을 높이고, 투명한 의사결정을 지원한 사례가 있다 (Kwon, 2023). 의료 분야에서는 딥러닝 기반의 무릎 골관절염 진단 모델에 XAI를 적용하여 진단 근거를 명확히 제시함으로써 의료진의 신뢰를 확보하고 진단 정확도를 향상시킨 사례가 있다 (Rafique과 Ali, 2024). 또한, 제조 분야에서는 XAI를 통해 수주량 변화의 주요 요인을 분석하고, 이를 기반으로 자원 배분 및 운영 최적화를 실현하여 비용 절감과 생산성 향상에 기여한 사례가 있다 (Jung과 Cheolhee, 2024).

이러한 연구사례들은 XAI가 단순히 예측 결과를 제공하는 데 그치지 않고, 그 결과를 해석하고 시각화함으로써 다양한 산업에서 실질적인 의사결정을 지원하는 도구로 자리 잡고 있음을 보여준다. XAI의 이러한 특성은 고속철도 수요 예측과 같은 대규모 교통 인프라 운영에서도 중요한 기여를 할 수 있다. 특히, KTX와 같은 고속철도망은 수요 변동성이 높고, 다양한 외부 요인의 영향을 받는 복잡한 시스템이므로, 예측 결과에 대한 명확한 해석과 신뢰성이 확보되지 않는다면, 정책 설계와 자원 배분의 효과성이 크게 저하될 수 있다. 따라서 XAI를 활용하여 AI 기반 수요 예측 결과를 해석 가능하게 만들고, 이를 시각화하여 운영 전략 수립과 정책 결정 과정에서 보다 신뢰도 높은 의사결정을 지원할 수 있다.

이후의 내용은 한국철도공사가 제공한 데이터와 전처리 과정을 소개하고, 승차인원수 예측을 위해 사용된 AI와 XAI 알고리즘의 소개, 그리고 마지막으로 2025년도 예측에 대한 연구결과를 제시하며 마무리한다.

# Methods

* 1. *Data Preprocessing*

본 연구에서는 경부선, 경전선, 동해선, 전라선, 호남선 총 5개의 월별 승차인원수 요를 예측하는 것이 목적이다. 한국철도공사 철도연구원으로부터 제공받은 2015년 1월부터 2024년 3월까지 약 10년간의 “수송-운행일-주운행” 그리고 “시종착역별 열차운행” 정보가 담긴 데이터베이스를 결합하여 2025년 12월까지의 월별 KTX 수송수요를 예측하는데 활용하였다. 그리고 수요예측에 도움이 될 수 있는 다양한 파생 변수들을 생성하였다. 첫째로, 과거의 수요가 현재 또는 미래의 수요에 영향을 줄 수 있기 때문에 과거 1개월~12개월 전의 수요를 “과거 승차인원수” 파생변수로 생성하였다. 둘째로, 버스나 자하철과 같은 대중교통과 달리 요일이나 이벤트에 따라 수요의 변화가 느리게 발생하기 때문에, 시계열데이터에서 각 월의 실제 날짜수, 주말수, 주중수, 공유일수, 명절 수 등의 “시간정보” 파생변수를 결합하였다. 셋째로, 대외적인 경제상황과 환경변화를 반영하기 위해서 한국의 주식 시장 지표와 소비자의 물가수준, 그리고 코로나 시기의 예방접종 인원수, 격리자수, 사망자수, 정부대응 지수 등의 “외부환경” 파생변수를 반영하여 정교함을 높였다. 마지막으로 한국철도공사에서 제공받은 공급좌석 정보, 열차 정보, 운행정보 등을 재계산하여 “좌석 및 운행정보” 변수로 반영하여 수요예측에 활용하였다. 표 1에 생성된 기본적인 파생변수들을 포함하여 수요예측에 사용한 변수들의 원천데이터의 예시를 정리하였다.

표 1. KTX 수요예측을 위해 사용한 변수명과 의미

|  |  |  |  |
| --- | --- | --- | --- |
| **특성** | **변수명** | **예시** | **정의** |
| **열차종** | **주운행선** | 경부선 | KTX 주요 5개 노선 경부/경전/동해/전라/호남선 |
| **과거**  **승차인원수** | **승차인원수\_Lag1** | 3,464,111.00 | 1개월 전 승차인원수 |
| **승차인원수\_Lag2** | 3,318,669.00 | 2개월 전 승차인원수 |
| **승차인원수\_Lag3** | 3,647,548.00 | 3개월 전 승차인원수 |
| **승차인원수\_Lag4** | 3,547,176.00 | 4개월 전 승차인원수 |
| **승차인원수\_Lag5** | 3,643,417.00 | 5개월 전 승차인원수 |
| **승차인원수\_Lag6** | 3,291,100.00 | 6개월 전 승차인원수 |
| **승차인원수\_Lag7** | 3,352,224.00 | 7개월 전 승차인원수 |
| **승차인원수\_Lag8** | 3,279,768.00 | 8개월 전 승차인원수 |
| **승차인원수\_Lag9** | 3,274,361.00 | 9개월 전 승차인원수 |
| **승차인원수\_Lag10** | 6,921,600.00 | 10개월 전 승차인원수 |
| **승차인원수\_Lag11** | 3,281,372.00 | 11개월 전 승차인원수 |
| **승차인원수\_Lag12** | 3,152,500.00 | 1년 전 승차인원수 |
| **시간정보** | **운행년월** | 45,352.00 | 해당 열차가 운행된 연도와 월 |
| **일수** | 31.00 | 해당 월에 포함된 날짜의 수 |
| **주말수** | 15.00 | 해당 월에 포함된 주말(금토일) 수 |
| **주중수** | 16.00 | 해당 월에 포함된 주중(월화수목) 수 |
| **공휴일수** | 1.00 | 해당 월에 포함된 대체휴일을 포함한 공휴일의 수 |
| **명절수** | - | 해당 월에 포함된 대체휴일을 포함한 명절의 수 |
| **좌석 및 운행정보** | **공급차량수** | 62,658.00 | 운행될 수 있는 총 열차의 수 |
| **공급좌석합계수** | 3,301,402.00 | 열차 내에 판매할 수 있는 좌석의 수 |
| **승차수입금액** | 113,298,033,576.00 | 판매된 좌석의 총금액 |
| **승차인원수** | 3,473,501.00 | 판매된 좌석의 수 |
| **승차연인거리** | 8,239,441,412.00 | 승객이 타고간 거리 |
| **좌석거리** | 13,516,340,607.00 | 공급좌석이 이동가능한 거리 |
| **1인당수입율** | 1,006,017.52 | "승차수입금액"을 "승차인원수"로 나눈 값으로, 1인당 평균 수입미 |
| **공급대비승차율** | 32.33 | "승차인원수"를 "공급좌석합계수"로 나눈 값으로, 좌석 판매의 집중도 |
| **운행대비고객이동** | 52.84 | "좌석거리"를 "승차연인거리"로 나눈 값으로, 승객의 이용 집중도 |
| **관광** | - | 관광으로 편성된 열차 수 |
| **일반** | 97.00 | 일반으로 편성된 열차 수 |
| **일반/관광** | 408.00 | 일반/관광으로 편성된 열차 수 |
| **대수송** | - | 대수송으로 편성된 열차 수 |
| **임시** | - | 임시로 편성된 열차 수 |
| **확정** | 505.00 | 확정으로 편성된 열차 수 |
| **시발역** | 155.00 | 열차가 출발하는 시발역 종류의 수 |
| **종착역** | 155.00 | 열차가 도착하는 종착역 종류의 수 |
| **시발종착역** | 248.00 | 열차가 운행하는 "시발역+종착역" 노선의 수 |
| **열차운행횟수** | 3,791.00 | 운행을 한 총 열차의 수 |
| **외부환경** | **Stringency Index** | 1,463.82 | 코로나 진행정도 지수 |
| **Government Response Index** | 1,775.99 | 정부의 코로나 대응정도 지수 |
| **International Movement Restrictions** | 62.00 | 국가간 이동 제한정도 지수 |
| **Death People** | 82,667.00 | 코로나 사망자 수 |
| **Vaccinated People** | 1,246,537,550.00 | 백신접종을 시작한 인원수 |
| **Fully Vaccinated People** | 1,014,704,524.00 | 백신접종이 완료된 인원수 |
| **Containment People** | 1,918.90 | 격리된 인원수 |
| **Confirmed People** | 10,586,338.00 | 코로나 확진자 수 |

예측 대상인 승차인원수는 종속변수로 사용되고 나머지 44개의 변수들은 독립변수로 사용된다. 분석기간은 모든 노선의 2015년 1월부터 2024년 3월까지이며, 그 중 시계열 순서로 2023년 4월부터 2024년 3월까지의 데이터는 검증(Validation)셋으로 나머지는 학습(Training)셋으로 사용된다. 그리고 최종 예측 목표인 2024년 4월부터 2025년 12월까지가 테스트(Test)셋이다. 데이터에 특별한 결측치는 존재하지 않으며 변수들의 값의 범위가 다양하기 때문에 동일하게 조정하기 위해 스케일링(Scaling)을 적용하였다. 마지막으로 전처리가 완료된 원천데이터를 기반으로 한국철도공사에서 중시하는 파생지표들로 변경한 후 최종적으로 모델링의 입력으로 활용된다

* 1. *Machine Learning Algorithm: Bagging and Boosting*

기본적으로 머신러닝 알고리즘은 예측 오차를 줄이는 방향으로 설계되었다. 오차는 bias and variance로 분리될 수 있는데, 예측의 안정성에 초점을 두어 variance를 줄이기 위해 샘플링 기법을 활용하는 Bagging과 성능에 초점을 두어 bias를 줄이기 위해 반복적인 모델링을 활용하는 Boosting으로 구분될 수 있다. Bagging은 여러 개의 샘플을 추출하여 각 샘플마다 모델링한 후 outputs의 투표로 최종 output의 label을 결정하는 방식이기에 bootstrap aggregating으로 불린다. 본 연구에 활용된 대표적 알고리즘은 Random Forest(RF)다. 변수의 중요도를 제공하긴 하지만 각 샘플별 모델링 과정에서 각 변수들의 우선순위를 평균한 것으로 positive or negative와 같은 영향력의 방향성을 포함하지는 못하는 단점이 있다.

Bagging처럼 Boosting도 classification and regression 문제 모두 활용할 수 있는 supervised learning이다. 본 연구에서는 XGBoost, LightGBM and CatBoost를 대표적인 boosting 알고리즘으로 활용하여 승차인원수 예측에 사용한다. 이 알고리즘들은 전체 데이터를 학습할 때 잘 학습되지 않은 에러들을 가중치를 높여 재학습을 하면서 성능을 향상시키는 방향으로 설계되었다.

* 1. *Deep Learning Algorithm: MLP, RNN*

딥러닝은 인공지능의 한 방법론으로, 인간의 두뇌 구조에서 영감을 받아 개발되었다. 데이터의 복잡한 패턴들을 학습하기 위해 연속된 layer를 중첩하여 변수들의 모든 상호작용을 포함하여 의미있는 규칙들을 학습해 내는데 강점이 있다. 이러한 구조의 기본이 되는 알고리즘으로 multilayer perceptron (MLP)가 있으며, 이미지나 시계열 등의 데이터도 학습해 낼 수 있도록 layer의 구조나 흐름을 개선하여 convolutional neural network (CNN)과 recurrent neural network (RNN) 등으로 확장되었다. neural network structure를 가진 모든 알고리즘들은 앞서 소개한 머신러닝의 알고리즘 처럼 activation function만 선택적으로 변경함으로써 classification and regression 문제 모두에 활용가능한 universal approximator이다. 모델의 구조가 복잡하기에 수많은 가중치들을 한꺼번에 추정해야 하는데 feed-forward and backpropagation을 사용하여 loss function을 최소화하기 위해 gradient descent optimizer를 사용하였다. 머신러닝과 마찬가지로 다양한 nonlinear patterns를 학습함으로써 기존의 알고리즘보다 성능이 대폭 향상되어 인간이 인지하기 어려운 것들도 학습해내는데 유용하다.

* 1. *RNN 기반 알고리즘’*

*2.5.1 LSTM(1997)*

LSTM(Long Short-Term Memory)은 Recurrent Neural Network(RNN)의 확장된 형태로 개발되었다. 기본적으로 RNN 기반 구조를 가지며, 순환 연결을 통해 이전 시점의 정보를 다음 시점으로 전달하는 특성을 가진다. 기존 RNN에서는 시간이 지남에 따라 기울기가 소실되거나 폭발하는 문제가 발생하여 장기적인 의존성을 학습하는 것이 어렵기 때문에, 이를 해결하기 위해 모델은 특별한 게이트 구조와 셀 상태(cell state) 메커니즘을 도입하였다 (Alex과 Alex, 2012).

핵심 구성 요소는 셀 상태와 세 가지 주요 게이트인 입력 게이트(input gate), 삭제 게이트(forget gate), 출력 게이트(output gate)이다. 입력 게이트는 새로운 정보를 셀 상태에 추가할지를 결정하고, 삭제 게이트는 기존 정보를 유지하거나 제거하는 역할을 하며, 출력 게이트는 최종적으로 출력할 정보를 결정하는 기능을 수행한다. 이러한 구조를 통해 기존 RNN과 비교하여 장기 의존성을 보다 효과적으로 학습할 수 있다.

이러한 구조 덕분에 해당 모델은 장기적인 패턴을 보다 효과적으로 학습할 수 있으며, 자연어 처리(NLP), 음성 및 영상 처리, 금융 데이터 분석, 생체 신호 분석, 수요 예측(demand forecasting)과 같은 시계열 분석 문제에서 활용되고, 특정한 시계열 데이터 처리에서 강력한 성능을 유지하고 다양한 최적화 기법과 결합하여 지속적으로 발전하고 있다.

*2.5.2 GRU(2014)*

Gated Recurrent Unit(GRU)는 LSTM(Long Short-Term Memory)의 경량화된 변형으로, 동일하게 기존 RNN이 가진 기울기 소실 및 폭발 문제를 해결하기 위해 설계되었으며, 순환 연결을 통해 시간에 따른 정보를 유지하고 전달하는 특성을 갖는다. LSTM은 게이트 구조(입력 게이트, 삭제 게이트, 출력 게이트)를 통해 장기 의존성 문제를 해결하는 반면, GRU는 비교적 단순화된 게이트 구조(업데이트 게이트와 리셋 게이트)를 채택하여 계산 효율성을 높였다. 별도의 메모리 셀을 사용하지 않고 은닉 상태를 직접 조절하며, 적은 파라미터 수로도 효과적인 학습이 가능하여 작은 데이터셋에서도 성능 저하 없이 빠르게 학습할 수 있다 (Kyunghyun 등, 2014).

결론적으로, 해당 모델은 LSTM보다 구조가 단순하고 연산량이 적어 학습 속도가 빠르며, 장기 의존성을 효과적으로 학습할 수 있는 업데이트 게이트와 불필요한 정보를 조절하는 리셋 게이트를 활용하는 구조적 특성을 가지고 있어 기계 번역, 감성 분석, 텍스트 요약 등의 자연어 처리(NLP) 분야에 활용되고, 수요 예측 및 심전도, 뇌파 분석과 같은 시계열 데이터 처리에서도 높은 성능을 보인다.

* 1. *Model Explainability: SHapley Additive exPlanations (SHAP)*

블랙박스와 같은 인공지능의 예측 결과를 설명하기 위한 알고리즘인 SHAP(Shapley Additive exPlanations)는 LIME(local interpretable model-agnostic explanations)과 셰플리 값 (Shapley Value)를 연결하여 KTX 승차인원수 수요 예측에 대한 변수들의 기여 정도와 방향을 설명해준다. LIME은 임의로 데이터의 값을 변화시켰을 때 모델의 예측 결과의 변화를 추정하며 관련성을 계산한다. 그리고 게임 이론을 기반으로 개발된 셰플리 값은 이러한 변수들의 기여 정도를 계산하는 지표로 볼 수 있다. 따라서 변수들이 모든 조합에서 실제 값의 입력을 통해 생성되는 예측값의 변화 정도로 기여도를 추정한다. SHAP를 사용하여 KTX 승차인원수 예측에 기여하는 변수들의 정도와 기여 방향을 알 수 있고 복잡한 인공지능 알고리즘의 신뢰성을 높인다.

* 1. *Evaluation Metrics*

KTX 승차인원수의 예측 성능을 확인하고 설명의 신뢰성을 높이기 위해 6개의 검증지표로 확인한다. 이들은 RMSE(Root Mean Squared Error), MSPE(Mean Squared Percentage Error), MAE(Mean Absolute Error), MAPE(Mean Absolute Percentage Error), MedAE(Median Absolute Error), MedAPE(Median Absolute Percentage Error)이다. 모든 검증지표들은 실제 승차인원수와 예측된 값의 차이를 방정식으로 표현한 것으로써, 예측 성능이 좋을수록 낮은 수치들이 나오도록 구성되어 있다. 따라서 모든 검증지표의 수치가 낮을수록 KTX 승차인원수를 잘 예측하는 모델이라 볼 수 있다.

단, 승차인원수는 경부선의 경우 백만이 넘는 큰 수치이기 때문에 예측이 어느 정도 잘 되더라도 검증지표는 수치가 크게 나타날 수도 있을 것이다. 따라서 최적 변수의 조합이나 모델 선택, 그리고 한국철도공사 측에 예측 설명력을 이해시키는 경우에는 MSPE, MAPE, MedAPE 3가지의 검증지표를 주로 사용할 것이다. 왜냐하면 이 3가지의 지표들은 실제 승차인원수 대비 상대적인 차이를 퍼센트로 표현하기 때문에 이해하기가 수월하기 때문이다.

# Results

이번 섹션에서는 다양한 파생변수를 포함하는 44개의 독립변수와 종속변수인 승차인원수를 모델에 학습시킨 후 2025년도의 승차인원수 수요 예측 결과를 제시한다. 실제 승차인원수가 정확하게 예측이 되어야 왜 그러한 수치가 나타나게 되었는지 변수들의 기여도 또는 관련성과 같은 설명력의 신뢰성이 높아질 것이다. 선형회귀분석과 같은 전통적 모델은 예측수치를 위한 변수들의 가중치가 쉽게 설명되지만 실제 미래를 예측할 경우 성능이 낮아서 설명만 될 뿐 실제 비즈니스에 활용되기는 어렵다. 사회과학 분야는 과거 설명력에 집중하느라 미래에 얼마나 신뢰할 수 있는 수치인지 다소 경시하는 경향이 있다. 따라서 머신러닝과 딥러닝 알고리즘으로 미래 예측 결과를 우선적으로 확인하여 모델링 성능을 확인 후 다음으로 SHAP 알고리즘을 통해 변수들의 승차인원수 예측 설명력을 확인한다. 데이터 준비와 전처리, 그리고 모델링과 검증의 모든 데이터분석 프로세스는 python 3.9.20 버전을 사용하였다. 그리고 머신러닝과 딥러닝 알고리즘은 sklearn 1.5.2과 tensorflow 2.17.0 버전을 사용하였다. 마지막으로 설명력 제공을 위한 SHAP 알고리즘은 0.46.0 버전이 사용되었다.

* 1. *Performance and Prediction Process*

2025년도 승차인원수의 예측 전에 과거데이터로 간주되는 학습데이터와 미래데이터로 간주되는 검증데이터를 사용하여 머신러닝 및 딥러닝 알고리즘 모델 예측 성능을 검증한다. 대표적인 머신러닝 알고리즘인 Random Forest, XGBoost, LightGBM, 그리고 CatBoost와 대표적인 딥러닝 알고리즘인 MLP, RNN, LSTM, 그리고 GRU 알고리즘의 성능을 검증하였다. 공정한 비교를 위해 알고리즘의 파라미터들은 모두 동일하게 사용하였다. 총 8가지의 알고리즘으로 모두 검증기간에 대해 예측을 한 후 검증기간의 예측성능이 높은 알고리즘을 선택하여 테스트기간동안 예측을 하였다. 그림 3과 표 4에 경부선 승차인원수에 대한 검증기간 상위 5종 예측 시각화와 검증지표를 내림차순으로 제시하였다. 그림 3 (a)에 따르면 학습에 활용되지 않은 검증기간에서의 승차인원수를 가장 정확하게 예측한 알고리즘은 LightGBM 알고리즘이다. 그리고 실제 LightGBM 알고리즘을 사용하여 예측을 할 때, 미래 발생가능한 시나리오를 생성하기 위해 독립변수들의 값을 특정 표준편차 범위 내로 랜덤하게 변화시켰다. 총 10000번의 예측을 통해 각 월별 예측값의 범위인 신뢰구간이 측정되며 그 평균값을 계산할 수가 있다. 이것이 그림 3 (b)에서 노란색 음영과 빨간색 점선으로 표시되는 것이다. 이러한 과정을 통해 예측값이 발생할 수 있는 범위를 알 수 있고 여러가지 미래 시나리오를 반영 및 설명하는 예측이 가능하다. 퍼센트 검증지표인 MSPE, MAPE, 그리고 MedAPE의 평균치 기준 2.49% 밖에 되지 않은 낮은 수준의 예측 오류를 보인다 (표 4).

|  |
| --- |
| (a) 예측성능 상위5종의 승차인원수 예측 비교 |
| (b) 최고 예측성능인 LightGBM 알고리즘의 예측평균과 신뢰구간 |

그림 10. 경부선의 검증기간(2023년 4월 ~ 2024년 3월)에서의 예측성능 상위 5종 승차인원수 예측 시각화

표 2. 경부선의 검증기간(2024년 4월 ~ 2024년 3월)에서의 예측 성능지표 6종 순위

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **Algorithm** | **RMSE** | **MSPE** | **MAE** | **MAPE** | **MedAE** | **MedAPE** | **MSPE+MAPE+MedAPE** |
| **LightGBM** | **183,310** | **0.30%** | **148,672** | **4.39%** | **99,094** | **2.79%** | **2.49%** |
| **RandomForest** | 200,483 | 0.35% | 174,521 | 5.13% | 181,347 | 5.49% | 3.66% |
| **CatBoost** | 201,252 | 0.35% | 180,439 | 5.27% | 189,825 | 5.41% | 3.68% |
| **XGBoost** | 235,083 | 0.47% | 220,373 | 6.44% | 240,598 | 7.03% | 4.65% |
| **RNN** | 257,769 | 0.56% | 254,829 | 7.44% | 266,759 | 7.73% | 5.24% |
| **MLP** | 369,436 | 1.08% | 283,462 | 8.07% | 255,093 | 7.40% | 5.52% |
| **GRU** | 301,891 | 0.78% | 297,093 | 8.68% | 300,684 | 8.81% | 6.09% |
| **LSTM** | 307,550 | 0.80% | 299,494 | 8.74% | 311,268 | 9.20% | 6.25% |

검증 성능이 가장 높은 LightGBM 알고리즘을 사용하여 2025년의 승차인원수를 예측한다. 단, 학습된 모델을 그대로 사용하여 테스트 기간을 예측하는 것이 아니라, 학습 및 검증 기간을 모두 사용하여 LightGBM 알고리즘을 재학습 한 후 테스트기간을 예측하였다. 왜냐하면 이미 학습이 잘 된 모델링이라고 하더라도 검증기간을 제외하고 바로 테스트기간을 예측할 수야 있지만 너무 먼 미래를 예측하는 것이기 때문에 성능도 떨어질 뿐 아니라 최근 데이터를 학습에 사용하지 못하는 단점이 있기 때문이다. 따라서 학습데이터로 모델링 후 검증기간을 예측한 것처럼 학습 및 검증데이터를 모두 사용하여 모델링 후 예측기간의 수치를 확인하는 시계열 모델링을 진행하였다. 그 결과 경부선의 경우, 과거 350만명 정도의 승차인원수가 코로나 이슈를 통해 감소하다 회복중에 있으며 2025년도 2024년보다 근소하게 상승할 것으로 나타나고 있다. 그리고 약 325만~350만명 사이의 승차인원수 수요가 예상된다 (그림 4).

|  |
| --- |
| (a) 학습, 검증, 테스트기간 전체의 실제 승차인원수와 예측치 |
| (b) 테스트 기간인 2024년 3월부터 2025년 전체의 월별 승차인원수 예측치 |

그림 11. 경부선의 전체기간(2015년 1월 ~ 2025년 12월)에서의 월별 승차인원수 예측 시각화

* 1. *Forecasting of 2025 Year*

위와 같은 프로세스를 통해 나머지 KTX 노선별 예측결과를 확인한다. 5개 노선의 예측 결과 LightGBM 알고리즘이 2개의 노선에서, XGBoost 알고리즘이 2개의 노선에서, 그리고 LSTM 알고리즘이 1개의 노선에서 최고 검증 예측성능을 보였다 (표 3). 전라선의 경우 딥러닝 알고리즘이 성능이 가장 좋았지만, 나머지 경부선, 경전선, 동해선과 호남선은 머신러닝 알고리즘이 예측력이 좋았다. 그리고 MSPE, MAPE, MedAPE 2가지 검증지표의 평균값을 기준으로 예측 오류는 2.49% ~ 6.89%의 낮은 범위를 보이고 있다. 즉, 머신러닝 또는 인공지능 알고리즘을 통해 KTX 승차인원수를 5% 전후의 낮은 오류의 예측성능을 발휘할 수 있다.

표 3. KTX 각 노선별 예측 검증성능 최고 알고리즘과 검증지표 비교

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **KTX** | **Algorithm** | **MSPE** | **MAPE** | **MedAPE** | **Average** |
| **경부선** | **LightGBM** | 0.30% | 4.39% | 2.79% | **2.49%** |
| **경전선** | **LightGBM** | 0.94% | 9.35% | 9.22% | **6.50%** |
| **동해선** | **XGBoost** | 0.91% | 9.18% | 8.40% | **6.16%** |
| **전라선** | **LSTM** | 0.99% | 9.77% | 9.92% | **6.89%** |
| **호남선** | **XGBoost** | 0.37% | 5.22% | 5.17% | **3.59%** |

그리고 경부선을 제외한 나머지 경전선, 동해선, 전라선, 그리고 호남선의 검증 및 테스트 예측 결과를 그림 5에 제시하였다. 모든 노선에서 머신러닝과 딥러닝 알고리즘이 전반적으로 승차인원수의 추세 뿐만 아니라 증가와 감소폭도 최대한 따라가기 위해 애쓰고 있음이 나타난다. 그리고 최종적으로 2025년의 KTX 승차인원수 수요는 경부선은 평균적으로 약 326만명, 경전선은 62만명, 동해선은 52만명, 전라선은 68만명, 그리고 호남선은 96만면 수준으로 예측되었다. 코로나 이전으로 볼 수 있는 약 5년 전인 2019년 대비 6.61~ 13.13% 수준으로 증가할 것으로 예상되며, 작년인 2024년 대비는 약 3.01 ~ 9.45% 수준의 감소가 예상되었다 .

|  |
| --- |
| (a) 경전선의 검증성능 및 예측 |
| (b) 동해선의 검증성능 및 예측 |
| (c) 전라선의 검증성능 및 예측 |
| (d) 호남선의 검증성능 및 예측 |

그림 12. 경전선, 동해선, 전라선, 호남선의 검증기간 및 테스트기간에서의 월별 승차인원수 예측 시각화

표 4. KTX 고속철도의 2025년도 승차인원수 예측 평균치와 코로나 전후의 과거 대비 증감율 비교

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| KTX | 2019년 | 2024년 | 2025년 | 증감율%(2019-2025) | 증감율%(2024-2025) |
| 경부선 | 3,000,064 | 3,418,760 | 3,261,319 | 8.71% | -4.61% |
| 경전선 | 552,302 | 675,125 | 624,792 | 13.13% | -7.46% |
| 동해선 | 470,231 | 573,080 | 521,482 | 10.90% | -9.00% |
| 전라선 | 580,489 | 759,476 | 687,733 | 18.47% | -9.45% |
| 호남선 | 905,066 | 994,880 | 964,904 | 6.61% | -3.01% |

* 1. *Explainability of 승차인원수 Forecating*

인공지능 알고리즘은 전통적인 선형회귀분석 형태의 알고리즘 보다 데이터의 복잡한 패턴과 변수들의 상호작용들을 모두 학습하기 때문에 구조가 복잡하여 승차인원수 예측에 기여하는 독립변수들의 기여도를 쉽게 확인하기 어렵다. 하지만 SHAP 알고리즘을 사용하면 미래 승차인원수 예측치가 어떤 독립변수들의 상대적 기여로 계산되었는지 추정할 수 있다. 따라서 2025년도 KTX 고속철도의 승차인원수 예측 뿐만 아니라 신뢰할 수 있는 설명력을 시각적으로 제공해준다.

SHAP는 특정 시점의 독립변수들의 샘플 수치들에 대해 승차인원수가 얼마일지 예측할 뿐만 아니라 각 독립변수들의 수치값과 기여 방향을 제시해준다. 이러한 결과를 누적한다면 승차인원수에 대한 독립변수들의 일반화된 기여 설명력을 확인할 수 있다. 단, 본 연구에 활용된 샘플 시점들 결과의 누적기준 일반화이기 때문에 샘플 시점이 향후 늘어난다면 해석이 변경될 수도 있다. 샘플의 양이 많아진다면 더욱 신뢰할 수 있는 설명력을 가질 것은 분명하다. 그림 5는 표 3에서 추정된 노선별 예측 최고성능 알고리즘을 기준으로 추정된 SHAP의 독립변수 기여도를 시각화 한 것이다.

(왼쪽)세로축은 승차인원수에 영향을 주는 독립변수의 우선순위를 내림차순으로 정렬한 것이다. 즉, 경부선을 기준으로 “좌석당단가, 좌석회전율, 공급차량수, 1열차당승차인원, 국가이동제한정도” 등의 순서로 승차인원수 예측에 기여를 하고 있다. (오른쪽)세로축은 변수들의 값이 낮은/높은 경우 파란색/빨간색 계열로 표시하여 값에 대응되는 승차인원수의 예측값의 변화를 표시하였다. 마지막으로 가로축에 승차인원수의 예측값을 표시하였다. SHAP 값이 0보다 작으면 부정적인 기여를 0보다 크면 긍정적인 기여를 의미한다. 각 시점마다 변수의 수치는 다양하게 분포할 수 있고 각 수치별 승차인원수에 대한 변수들의 기여 방향도 얼마든지 변화될 수 있다. 따라서 변수들의 값과 대응되는 SHAP 예측치를 모두 누적하여 표현하면 변수 값의 변화에 따른 승차인원수 기여 방향을 확인할 수 있다.

경부선의 “1좌석당단가”는 값이 작을때(파란색계열) 승차인원수의 부정 기여에 많이 분포되어 있고 값이 커지면(빨간색계열) 승차인원수의 긍정 기여에 많이 분포 되어 있다. 따라서 1좌석당단가가 비싸질수록 승차인원수를 높이는 영향을 준다. 반대로 코로나가 발생하여 “국가이동제한정도”가 증가하면 승차인원수를 낮추는데 기여한다. 따라서 미래의 승차인원수에 증가에 가장 기여를 많이 하는 변수의 상위 3가지 변수는 "1좌석당단가, 좌석회전율, 공급차량수"이며 승차인원수 감소에 가장 기여를 많이하는 변수 상위3가지는 "국가이동제한정도, 코로나진행정도, 승차인원수\_lag3"로 볼 수 있다. 그런데 이러한 변수들의 기여정도는 다른 노선에선 다르게 나타난다. 경전선의 경우 승차인원수를 증가시키는 변수 상위 3가지는 경부선과 달리 “승차인원수\_lag1, 1열차당승차인원, 열차운행횟수”로 나타났다. 반면 승차인원수를 감소시키는 변수 상위 3가지는 “승차인원수\_lag12, 격리된자수, 1인당단가”가 해당된다. 마찬가지의 해석을 다른 노선들에서도 가능하지만, 중요한 것은 특정 변수가 모든 노선에서 긍정 또는 부정적 기여를 한다고 일반화 하는것은 위험하다는 것이다. 따라서 더욱 실제 비즈니스에 활용하기 위해선 실시간으로 정량적인 예측과 설명력을 빠르게 추정하여 활용하는 것이 훨씬 용이하다. 그러한 목적을 위해 본 연구는 누구나 현업에서 빠르고 쉽게 사용할 수 있는 설명가능한 인공지능을 활용한 비즈니스 애널리틱스 방법론을 제시하였다. 승차인원수를 높은 성능으로 미리 예측할 뿐 아니라 그 원인을 설명함으로써 비즈니스의 활용도를 높일 수 있다.

|  |  |
| --- | --- |
| (a) 경부선 | (b) 경전선 |
| (c) 동해선 | (d) 전라선 |
| (e) 호남선 | |

그림 13. KTX 노선별 승차인원수 예측에 대한 변수들의 기여 설명력 기반 의사결정

# Conclusions

현재 25개국 이상이 고속철도를 도입하여 주요 교통수단으로 활용하고 있으며, 이러한 국가는 고속철도를 통해 물류와 인구의 이동성을 극대화하며 교통체계의 효율성과 환경적 지속 가능성을 동시에 달성하고 있다. 특히, 고속철도는 전기차보다 12배, 내연기관 차량보다 26배 낮은 CO2 배출량을 기록하며, 지속 가능한 교통수단의 대표 사례로 자리 잡고 있다. 국토교통부는 미래 교통 시스템의 효율성을 제고하기 위해 수요 예측과 연계된 미래 선도 기술 개발을 적극적으로 추진하고 있다. 이러한 맥락에서, 고속철도의 수요 변화를 정밀하게 예측하는 것은 단순한 운영 효율성 개선을 넘어, 장기적인 교통 정책 수립과 지속 가능한 인프라 구축을 위한 핵심 요소로 작용한다. 효과적인 수요 예측을 위해서는 고속철도 이용 패턴을 정확히 반영할 수 있는 정교한 분석 기법이 필수적이다.

본 연구는 고성능 인공지능 알고리즘과 설명 가능한 인공지능(XAI, Explainable AI)을 활용하여 2025년도 KTX 수요를 정밀하게 예측하고, 예측 결과를 바탕으로 실질적인 비즈니스 및 정책적 의사결정을 지원하는 것을 주된 목적으로 3가지 항목에서 기여하고 있다. 첫째, AI 알고리즘을 활용하여 KTX 수요 예측의 정확도를 대폭 향상시킨다. 기존의 통계 기반 모델이 가지는 예측 성능의 한계를 극복하고, 고성능 알고리즘을 통해 비선형적 복잡한 데이터 패턴을 학습함으로써 더욱 현실성 높은 수요를 예측한다. 이를 통해, KTX 수요 변화의 복잡한 양상을 효과적으로 예측하고, 미래 교통 계획 수립에 기여할 것이다. 둘째, XAI 기술을 통해 높은 미래 예측 성능 뿐만 아니라 설명력을 동시에 제공함으로써 신뢰성을 높인다. SHAP(Shapley Additive Explanations) 알고리즘을 활용하여 예측 결과와 변수들의 기여 정도를 시각화 함으로써, 비즈니스적 정책 설계 및 자원 배분 의사결정 과정에서 신뢰도를 높인다. 셋째, 고성능 컴퓨팅 환경 없이도 실행 가능한 대표적인 머신러닝과 딥러닝 알고리즘을 활용하여, 한국철도공사 운영, 관리 및 정책 담당자가 현업에서 쉽고 간편하게 수요 예측 결과를 확인하고 실시간으로 의사결정을 내릴 수 있도록 활용도를 강화하였다.

결론적으로 본 연구는 AI 기반 수요 예측의 높은 정확성과 설명력을 통해 2025년도 KTX 수요 예측의 신뢰성과 실효성을 극대화하였으며, 데이터 기반 의사결정이 보다 효율적으로 활용가능한 방식으로 이루어져 지속 가능한 교통체계 구축에 도움이 될 것이다. 향후 다양한 교통 체계의 수요예측에도 활용가능한 연구 방향을 제시함으로써 교통 전반의 데이터기반 효율성을 높이는 데 중요한 학문적·실무적 시사점을 제공할 것이다. 또한 한국철도공사의 도메인 지식과 추가적인 고성능 AI 알고리즘을 활용하여 더욱 정교한 수요예측을 달성하는 것을 향후 과제로 남겨둔다.

# References

Alex, G., Alex, G. (2012). Long short-term memory. <Supervised sequence labelling with recurrent neural networks>, 37-45.

David, S., Valentin, F., Jan, G., Tim, J. (2020). DeepAR: Probabilistic forecasting with autoregressive recurrent networks. <International journal of forecasting>, **36**, 1181-1191.

Kyunghyun, C., Merriënboer,, V., Caglar, G., Dzmitry, B., Fethi, B., Holger, S., Yoshua, B. (2014). Learning phrase representations using RNN encoder-decoder for statistical machine translation. <arXiv preprint arXiv:1406.1078>, .

Shaojie, B., J, K., Vladlen, K. (2018). An empirical evaluation of generic convolutional and recurrent networks for sequence modeling. <arXiv preprint arXiv:1803.01271>, .

Shiyu, C., Yang, Z., Wei, H., Mo, Y., Xiaoxiao, G., Wei, T., Xiaodong, C., Michael, W., Mark, H. J., Thomas, H. (2017). Dilated recurrent neural networks. <Advances in neural information processing systems>, **30**, .

Union, I. (2023). HIGH-SPEED RAIL, the right speed for our planet - Key Messages. <11th UIC HS Congress>, **1**, 24.

기술과가치 (2020). 국토교통 2050 미래기술 도출을 위한 조사분석 연구. <국토교통과학기술진흥원>, .

Abhimanyu, D., Weihao, K., Andrew, L., Shaan, M., Rajat, S. and Rose, Y. (2023). Long-term forecasting with tide: Time-series dense encoder. *arXiv preprint arXiv:2304.08424*, .

Boris, O., Dmitri, C., Nicolas, C. and Yoshua, B. (2019). N-BEATS: Neural basis expansion analysis for interpretable time series forecasting. *arXiv preprint arXiv:1905.10437*, .

Cha, H., Oh, Y., Song, J. and Lee, T. (2019). KTX passenger demand forecast with multiple intervention seasonal ARIMA models. *응용통계연구*, **32**, 139-148.

Cristian, C., Kin, O., Boris, O., Federico, R., Max, C. and Artur, D. (2023). Nhits: Neural hierarchical interpolation for time series forecasting. *Proceedings of the AAAI conference on artificial intelligence*, **37**, 6989-6997.

David, G., Mark, S., Jaesik, C., Timothy, M., Simone, S. and Guang-Zhong, Y. (2019). XAI—Explainable artificial intelligence. *Science robotics*, **4**, eaay7120.

Jang, J. H. and Kim, M. K. (2022). An Explanatory Analysis of Electricity Demand Prediction Using XAI and N-BEATS Models. *대한전기학회 학술대회 논문집*, 277-278.

Jeong, H. and Lim, C. (2019). A review of articial intelligence based demand forecasting techniques. *응용통계연구*, **32**, 795-835.

Jung, Y. and Cheolhee, Y. (2024). A Study on Forecasting Order Quantity from Manufacturing Supply Chain Data using XAI. *Journal of Korean Institute of Information Technology*, **22**, 41-53.

Kim, K. and Kim, H. (2011). KTX Passenger Demand Forecast with Intervention ARIMA Model. *Journal of the Korean Society for Railway*, **14**, 470-476.

Kin, O., Cristian, C., Grzegorz, M., Rafał, W. and Artur, D. (2023). Neural basis expansion analysis with exogenous variables: Forecasting electricity prices with NBEATSx. *International Journal of Forecasting*, **39**, 884-900.

Kwon, J. (2023). A Study on the Applicability of eXplainable Artificial Intelligence(XAI) Methodology by Industrial District. *Global Business Administration Review*, **20**, 195-208.

Rafique, A. and Ali, I. (2024). Knee Osteoarthritis Analysis Using Deep Learning and XAI on X-rays. *IEEE Access*, .

Shanlang Lin, Prithvi Raj Dhakal, and Zhaowei Wu. 2021. “The Impact of High-Speed Railway on China’s Regional Economic Growth Based on the Perspective of Regional Heterogeneity of Quality of Place.” *Sustainability* 13, no. 9: 4820. <https://doi.org/10.3390/su13094820>.

Sima, S. N., Neda, T. and Akbar, N. (2018). A comparison of ARIMA and LSTM in forecasting time series. *2018 17th IEEE international conference on machine learning and applications (ICMLA)*, 1394-1401.

Syama, R., Jan, G., Lorenzo, S., Valentin, F., David, S., Yuyang, W. and Tim, J. (2023). Deep Non-Parametric Time Series Forecaster. *arXiv preprint arXiv:2312.14657*, .

국토교통부. 2024. "2024년 고속철도 연간 이용객 1억 1,658만 명 기록, 국민의 대표 이동수단으로 자리 잡아." *국토교통부 보도자료*, 1월 16일. <https://www.molit.go.kr/USR/NEWS/m_71/dtl.jsp?lcmspage=1&id=95090606>.

심진호, 오윤식, 오영택 and 금수희 (2024). A study of high-speed railway demand forecasting using nonlinear data pattern learning. *한국철도학회 학술발표대회논문집*, 114-114.

유현아, 남기찬, 홍사흠, 정동호, 김현중, 이용상, 최진석. 2024. *고속철도 20년, 국토균형발전 효과분석과 향후 과제*. 국토연구원.