|  |
| --- |
| **고성능 인공지능 알고리즘 활용**  **2025년도 KTX 수송수요 예측 연구** |

**- 자문보고서 -**

2024. 06.



**제 출 문**

**한국철도공사 연구원장 귀하**

**본 보고서를**

**「** **고성능 인공지능 알고리즘 활용 2025년도 KTX 수송수요 예측 연구 」**

**자문 최종보고서로 제출합니다.**

2024. 06.

**인천대학교 글로벌정경대학 교수 김 경 원**

목차

1. 연구 개요

1-1. 연구의 배경 및 목적

1-2. 연구의 대상 및 방향

2. 수요예측 데이터

2-1. 기술적 분석

2-2. 데이터 전처리

3. 예측 방법론

3-1. 머신러닝 기반 예측

3-2. 딥러닝 기반 예측

4. 수송수요 예측 결과

4-1. 2024년도 예측 성능 검증

4-2. 2025년도 예측 결과

5. 결론

5-1. KTX 수요예측 결론

5-2. 향후 연구 방향

# 연구 개요

## 1.1연구의 배경 및 목적

### 연구의 배경

○ KTX 수송수요(승차인원수)를 정확하게 예측하기 위해 고성능의 예측 모델의 개발이 필요함

- 최근 각광받는 인공지능 알고리즘을 활용하여 수송수요 예측 수행

- 다양한 형태의 인공지능 알고리즘을 활용하여 시계열 데이터인 수송수요에 적용하여 KTX 노선별 예측 수행

### 1.1.2 연구의 목적

○ 2025년도 KTX 수송계획 및 열차운행계획 수립을 위해 필요한 2024년도 단기 및 2025년도 중기 수송수요예측을 위한 고성능 인공지능 모델을 활용한 수송수요를 예측 방법을 구축하는 것이 주 목적으로 함

○ 기존 예측 방법론의 고도화를 달성하기 위해서, 시계열 데이터인 KTX 노선별 수송수요 대상 머신러닝 및 딥러닝 기반의 전통적인 알고리즘부터 최신 알고리즘을 활용하여 예측 분석을 진행함

- KTX 수요에 영향을 줄 수 있는 다양한 파생변수들을 생성하여 모델에 반영함으로써 향후 더욱 복합적인 변수들의 반영을 쉽게 하여 성능 향상에 쉽게 기여할 수 있도록 개발

- 머신러닝과 딥러닝은 과거부터 현재까지 수요예측 성능 향상에 기여할 수 있도록 빠르게 진화하였기 때문에 이를 반영하여 개발

- KTX 노선별 주중, 주말, 그리고 전체 수송수요 예측성능을 검증하고 가장 성능이 높은 모델을 중심으로 2025년도 수송수요를 예측하도록 개발

## 1.2연구의 대상 및 방향

### 1.2.1 연구의 대상

○ KTX 수송수요 예측을 위해 총 5개의 KTX 노선을 대상으로 함

- 경부선, 경전선, 동해선, 전라선, 호남선

- 각 노선은 수송수요의 차이가 있으므로 노선 별로 구분하여 예측

○ 수송수요 예측은 주말, 주중, 그리고 전체를 대상으로 함

- 주말의 기준은 “금토일”이며, 주중의 기준은 “월화수목”으로 함

- 각 요일에 따라서 수송수요의 차이가 있기 때문에 구분하여 예측

○ 수송수요 예측의 시간 빈도는 “월(Month)”을 기준으로 함

- 2025년도 수송수요 예측을 위해 일별과 월별 데이터를 활용할 수 있는데, 본 연구에서는 2025년도 1월부터 12월까지의 수요가 필요하기 때문에 입력과 출력에 사용되는 데이터의 시간은 월을 사용함

○ 각 노선별 주말/주중/전체로 구분하여 총 15개의 월별 KTX 수송수요 예측

|  |  |  |  |
| --- | --- | --- | --- |
| **시간빈도** | **노선** | **대상** | **총 15개 월별 KTX 수송수요** |
| 월별 | 경부선 | 전체 | 월별 경부선 전체 수요 |
| 주말 | 월별 경부선 주말 수요 |
| 주중 | 월별 경부선 주중 수요 |
| 경전선 | 전체 | 월별 경전선 전체 수요 |
| 주말 | 월별 경전선 주말 수요 |
| 주중 | 월별 경전선 주중 수요 |
| 동해선 | 전체 | 월별 동해선 전체 수요 |
| 주말 | 월별 동해선 주말 수요 |
| 주중 | 월별 동해선 주중 수요 |
| 전라선 | 전체 | 월별 전라선 전체 수요 |
| 주말 | 월별 전라선 주말 수요 |
| 주중 | 월별 전라선 주중 수요 |
| 호남선 | 전체 | 월별 호남선 전체 수요 |
| 주말 | 월별 호남선 주말 수요 |
| 주중 | 월별 호남선 주중 수요 |

<표> 1‑1. 본 연구의 분석 대상인 총 15개의 월별 KTX 수송수요

### 1.2.2 연구의 방향

○ **[데이터 전처리]** KTX 수송수요 예측에 실제 추가 변수들이 도움이 되는지 여부를 확인하고 필요시 반영을 위해서 다양한 파생 또는 외부 변수 후보들을 준비함

- 파생변수(Derived Variables)란, 기존 정보나 변수를 활용하여 재계산하거나 조합하여 새롭게 추가 반영된 변수를 의미

- 시간정보를 포함하고 있는 시계열 데이터를 통해 각 월의 실제 날짜 수, 주말 수, 주중 수, 공휴일 수, 명절 수 등의 파생변수 생성함

- COVID-19 종료 이벤트를 반영하기 위해 예방접종 인원수, 격지자 수, 사망자 수, 정부대응 지수 등의 파생변수를 결합함

- 경제상황의 변동을 모델에 반영하기 위해서 한국의 주식 시장 지표와 소비자의 물가 수준과 관련이 있는 지표들을 결합함

- 그 외에 제공받은 수송수요를 포함한 공급좌석 정보, 열차 정보, 운행 정보 등을 재계산하여 파생변수들을 생성함

○ **[데이터 분리]** 수송수요 예측 모델링에 사용되는 데이터는 크게 학습(Training), 검증(Validate), 예측(Test) 3가지로 구분되어 활용됨

- 모델링에 사용되는 데이터를 학습(Training)이라고 하며, 모델링 완료 후 학습에 사용되지 않은 데이터로 예측 성능을 검증하는 데 이를 검증(Validate)이라고 함, 그리고 검증과정에서 성능이 높은 모델을 사용하여 필요한 기간의 수송수요를 예측하는 것을 예측(Test)라고 함

- 학습과 검증, 그리고 예측에 사용되는 데이터의 기간은 아래와 같음

- 학습(Training) 기간: 2015년 1월 ~ 2023년 3월

- 성능(Validate) 기간: 2023년 4월 ~ 2024년 3월

- 예측(Test) 기간: 2024년 4월 ~ 2025년 12월

○ **[모델링]** 수송수요 예측에 활용되는 모델은 머신러닝(Machine Learning) 및 딥러닝(Deep Learning) 기반으로 시계열 수송수요를 분석하여 성능을 비교함

- 머신러닝 알고리즘은 딥러닝 알고리즘보다 성능은 낮은 편으로 알려져 있지만 학습과 예측에 소요되는 시간이 매우 빠르고, 데이터의 품질이 좋은 경우 딥러닝에 견주어도 성능이 떨어지지 않는 알고리즘으로 수송수요 예측 성능 비교에 활용되는 베이스라인(Baseline) 모델로 활용됨

- 딥러닝은 데이터의 종류에 따라서 활용될 수 있는 알고리즘이 개발되고 진화되고 있으며, 시계열 데이터에 활용될 수 있는 알고리즘은 다층 퍼셉트론(Multi-layer Perceptron, MLP), 순환신경망(Recurrent Neural Network, RNN), 그리고 최근 생성형AI에 활용되고 있는 트랜스포머(Transformer)가 있음

- 월별 데이터의 특성상 데이터의 개수가 많지 않아서 트랜스포머를 제외한 딥러닝의 MLP 및 RNN 계열에서 최근까지 개발된 알고리즘들을 비교 모델로 활용함