



고성능 시계열 딥러닝 알고리즘을 활용한 미래 교통수요 예측

(Forecasting Future Transportation Demand with High-Performance Time-Series Deep Learning)

차명주¹, 오영택², 이승연², 김경원^{1,*}

¹국립대학법인 인천대학교
글로벌경영대학 무역학부

²한국철도공사 철도연구원

CONTENTS

- I. 연구배경 및 목적
- II. 분석 대상 및 데이터처리
- III. 인공지능 이해 및 예측 방법론
- IV. 성능검증 및 예측결과
- V. 결론 및 향후 방향

CONTENTS

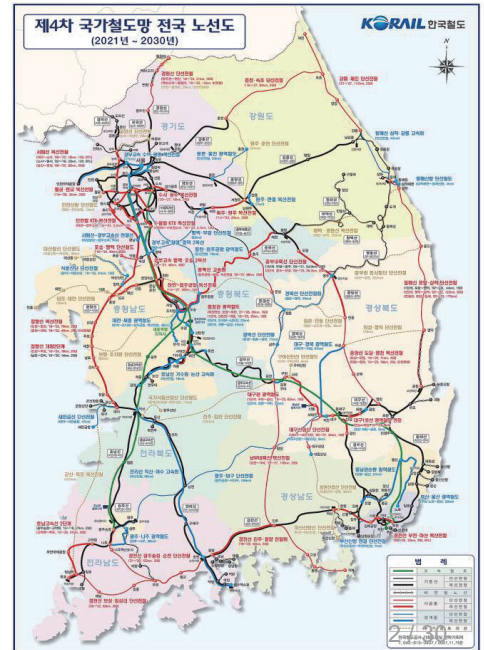
- I. 연구배경 및 목적
- II. 분석 대상 및 데이터처리
 - 분석대상(Target)
 - 데이터이해(Descriptive Statistics & Visualization)
 - 데이터처리(Data Preprocessing)
- III. 인공지능 이해 및 예측 방법론
 - 모델링 진화와 작동방식의 이해
 - 머신러닝 기반 알고리즘
 - 딥러닝 기반 알고리즘
- IV. 성능검증 및 예측결과
 - 2024년도 노선별 예측검증 방향
 - 2024년도 노선별 예측검증 요약
 - 2024년도 노선별 예측검증 및 2025년도 예측
 - 과거년도 대비 2025년도 예측치 일평균 비교
- V. 결론 및 향후 방향

I. 연구배경 및 목적

➤ 배경

▪ 고속철도의 사회적 & 경제적 기여

- ① 고속철도는 수도권과 지방 주요 도시를 연결하는 중요한 교통 인프라로서, 수도권에 집중된 경제 활동을 지방으로 분산시킴으로써 국토의 균형 발전을 촉진
- ② 제4차 국가철도망 구축계획은 고속철도 노선을 전국적으로 확충하고, 기존 노선의 운영 효율을 극대화함으로써 대한민국 전역의 교통망을 더욱 촘촘하게 연결하고, 지방의 교통 접근성을 개선하려는 장기 계획



Kyungwon Kim

INCHEON NATIONAL UNIVERSITY

I. 연구배경 및 목적

➤ 배경

▪ 고속철도의 사회적 & 경제적 기여

▪ 고속철도 수요의 증가와 변화

- ① 고속철도는 수도권과 지방 주요 도시를 연결하는 중요한 교통 인프라로서, 수도권에 집중된 경제 활동을 지방으로 분산시킴으로써 국토의 균형 발전을 촉진
- ② 한국의 경제 성장률은 고속철도 수송수요에 중요한 영향을 미치는 요인 중 하나임. 경제가 성장하면서 가처분 소득이 증가하고, 이에 따라 비즈니스 및 여가 목적으로 고속철도를 이용하는 수요 증가
- ③ 수도권으로의 인구 집중이 계속되면서 수도권과 지방 간 교통 수요는 꾸준히 증가
- ④ 고령 인구가 증가함에 따라 편리하고 안전한 교통수단을 선호하는 경향이 높아지고 있으며, 이에 따라 도로 교통보다는 고속철도를 이용하려는 수요가 늘어나는 추세임
- ⑤ 원격 근무와 비대면 업무가 보편화되면서 평일 출퇴근 수요는 일부 감소했지만, 반대로 관광 및 레저 수요가 증가하는 경향
- ⑥ 주말이나 공휴일에 주요 관광지로 이동하는 수요가 늘어나고 있어 고속철도 운영에 변화를 요구

Kyungwon Kim

INCHEON NATIONAL UNIVERSITY

3 / 30

I. 연구배경 및 목적

➤ 배경

- 고속철도의 사회적 & 경제적 기여
 - 고속철도 수요의 증가와 변화
 - 고속철도 수요예측의 필요성
- ① 고속철도의 수송수요는 경제 성장, 인구 구조 변화, 사회적 환경 변화, 관광 및 여가 활동 증가 등 다양한 요인에 의해 영향을 받으며, 이러한 수요 변화를 정확히 예측하는 것은 철도 운영 효율성 향상에 필수적
 - ② 철도 운영자에게 운영 계획을 수립하고 자원을 효율적으로 배분하는 데 필요한 정보를 제공함으로써, 전체적인 운영 효율을 극대화할 수 있도록 돕는 중요한 과정
 - ③ 수도권과 비수도권 간의 경제적 격차를 해소하고, 지방의 교통 접근성을 높이기 위해 비수도권 지역의 고속철도 수송수요를 정확히 파악하여 필요한 지역에 적절한 교통 인프라를 제공하는 것이 중요
 - ④ 철도 운영을 효율적으로 계획하면, 도로 교통에서 발생하는 차량 혼잡을 줄이고 탄소 배출 절감에 기여
 - ⑤ 철도 운영자는 예측된 수요를 바탕으로 열차 편성과 배차 계획을 최적화하여 승객이 물리는 시간대의 혼잡도를 낮추고 더욱 쾌적하고 편리한 이동이 가능

I. 연구배경 및 목적

➤ 목적

- ① 2025년도 고속철도 수요를 예측하여, 철도공사의 수송계획 및 열차 운행계획 수립의 기초 자료 제공
- ② 정확한 수송수요 예측을 통한 각 지역별 수송 목표와 수익 목표를 효과적으로 배분
- ③ 예측된 2025년 수송수요를 기반으로 열차의 운행 빈도와 배차 간격을 최적화하여 승객의 편의성 증대
- ④ SR의 경전선, 동해선, 전라선 운행 확대로 인해 기존의 수송 패턴에 변화가 예상되기에 이러한 외부 요인을 반영하여 보다 정확한 수요 예측을 수행
- ⑤ 수도권과 지방 간 수송수요의 격차를 분석하고, 이에 맞는 교통 인프라 배치 계획을 수립

➤ 연구의 범위

- 2015년 1월부터 2024년 6월까지의 월별 수송 실적을 기반으로 2024년도 및 2025년도 수송수요예측
- 대한민국의 주요 고속철도 노선인 경부선(서울~부산), 호남선(서울~광주), 전라선(서울~여수엑스포), 경전선(서울~진주), 동해선(서울~강릉) 등 총 5개 노선
- 기존 예측 방법론의 고도화를 위해 시계열 데이터에 적용가능한 고성능 AI로 학습된 전문가로 예측
- 미래 예측된 시계열 데이터와 추가 변수를 활용하여 모델의 정밀도를 높임

CONTENTS

I. 연구배경 및 목적

II. 분석 대상 및 데이터처리

- 분석대상(Target)
- 데이터이해(Descriptive Statistics & Visualization)
- 데이터처리(Data Preprocessing)

III. 인공지능 이해 및 예측 방법론

- 모델링 진화와 작동방식의 이해
- 머신러닝 기반 알고리즘
- 딥러닝 기반 알고리즘

IV. 성능검증 및 예측결과

- 2024년도 노선별 예측검증 방향
- 2024년도 노선별 예측검증 요약
- 2024년도 노선별 예측검증 및 2025년도 예측
- 과거년도 대비 2025년도 예측치 일평균 비교

V. 결론 및 향후 방향

II. 분석대상 및 데이터처리

➤ 분석대상(Target)

- 예측기간: 2025년 1월 ~ 2025년 12월
- 대상노선: 경부선, 경전선, 동해선, 전라선, 호남선 5종
- 데이터빈도: 월(Month)
- 대상요일: 전체, 주말(금토일), 주중(월화수목)

→ 2025년 12월까지의 5개 노선 월별 전체, 주말과 주중 15종을 독립적으로 구분하여 수요예측

데이터 변수		설명
종속변수(Y)		예측 대상인 승차인원수
독립변수(X) 36종	과거승차인원수	과거 1개월~12개월 전 승차인원수 데이터를 변수로 생성
	시간정보	양력/음력 기반 공휴일과 대체휴일로 주말수/주중수/공휴일 수 등 생성
	좌석 및 운행정보 (비공식)	공급차량수, 공급좌석합계수, 승차인원수, 1인당단가, 1인당 거리, 1좌석당단가, 좌석회전율, 1키로당단가, 승차율, 관광, 일반, 대수송, 임시, 확정, 시발역, 종착역, 시발종착역, 열차 운행횟수, 1열차당승차인원 등
	외부환경 (코로나 이슈 제외)	코로나 관련 정부기관 대응정도와 감염/격리/사망자 수 포함 경제환경 데이터 결합
데이터 분리	기간	의미
학습(Training)	2015년 1월 ~ 2023년 3월	모델 구축을 위해 사용되는 데이터
검증(Validate)	2023년 4월 ~ 2024년 3월	모델 구축 후 예측성능 검증에 사용되는 데이터
예측(Test)	2024년 4월 ~ 2025년 12월	검증성능이 높은 모델로 실제 예측에 사용되는 데이터

II. 분석대상 및 데이터처리

➤ 데이터이해(Descriptive Statistics & Visualization)

- 장기간 월별 전체/주말/주중 승차인원수 중심값(평균), 퍼진정도(표준편차), 최대기록(최대값) 확인

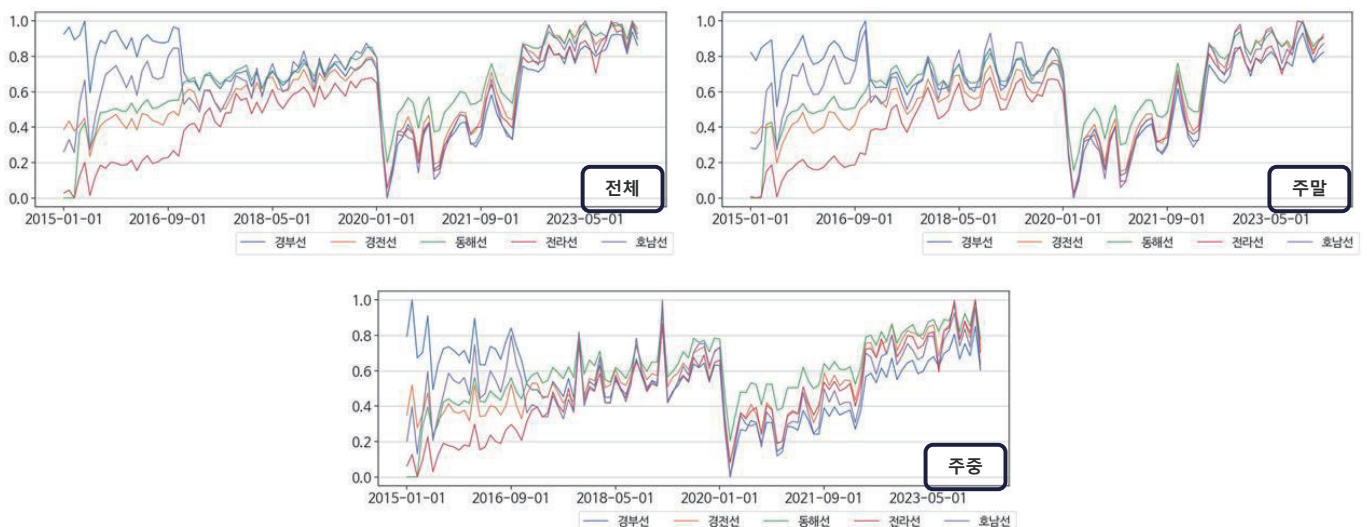
주운행선	전체주중주말	월별 승차인원수			일별 승차인원수		
		평균	표준편차	최대값	평균	표준편차	최대값
경부선	전체	2,387,398	1,210,415	3,894,486	78,431	39,765	127,942
	주말	1,207,555	631,933	2,144,330	92,565	48,441	164,374
	주중	1,179,843	596,738	1,968,430	67,831	34,307	113,168
경전선	전체	412,924	210,629	716,861	13,565	6,920	23,550
	주말	209,397	111,073	412,828	16,051	8,514	31,645
	주중	203,527	102,553	379,607	11,701	5,896	21,824
동해선	전체	109,223	218,167	612,267	3,588	7,167	20,114
	주말	52,069	104,835	324,842	3,991	8,036	24,901
	주중	57,154	114,084	345,258	3,286	6,559	19,849
전라선	전체	417,882	234,184	840,753	13,728	7,693	27,621
	주말	214,397	122,773	461,647	16,435	9,411	35,388
	주중	203,485	114,576	413,158	11,699	6,587	23,753
호남선	전체	675,680	331,543	1,088,042	22,198	10,892	35,745
	주말	339,963	173,952	583,089	26,060	13,334	44,697
	주중	335,716	163,390	588,227	19,301	9,394	33,818

- 월별 **경부선(약238만)**은 나머지 노선들의 합(약161만) 보다 많은 승객이 이용하고 있음
- 일평균 기준 주중보다 주말에 승차인원수가 많고 **주말 > 전체 > 주중** 순서로 **정확한 예측치 필요**
- 승차인원수 표준편차가 평균의 약50% 최대값은 평균의 약300%로 **높은 변동성을 담을 파생변수 필요**

II. 분석대상 및 데이터처리

➤ 데이터이해(Descriptive Statistics & Visualization)

- 장기간 월별 전체/주말/주중 승차인원수 차이를 **시각화**로 확인



- 증가/감소 폭이 과거에는 차이가 있었지만(Decoupling) **최근 움직임이 매우 유사한 동조화(Recoupling)**
- 10개 케이스 중 **특정 알고리즘이 다수 예측에서 우월할 수 있으며 변동반영을 위한 다양한 파생변수 필요**

II. 분석대상 및 데이터처리

➤ 데이터처리(Data Preprocessing)

(1) 데이터결합: 주은행선 좌석정보 데이터 + 시종착역별 운행정보 데이터 + 변수일원화

✓ 데이터 전처리: (0) 쓸모 없을 뻔한 Raw를 쓸모 있는 Data로 변환

100	T50	횡수	111	TPU	...
few	Gds	Hvi	Rew	Fa	...
Fre	CT	QTP	D	합	...
'1'	1	23	22	NaN	43
76	NaN	43	32	1	8
'Hi'	NaN	NaN	NaN	NaN	87
23	98	NaN	64	46	NaN
90	NaN	'WW'	24	'KK'	4
NaN	2	NaN	NaN	NaN	6
64	NaN	90	'IU'	4	76

빈도	시간	총량	기간	누적	...
1	1	23	22	21	43
76	33.3	43	32	1	8
5	33.3	52	35	21	87
23	98	52	64	46	61
90	33.3	2	24	33	4
55	2	52	35	21	6
64	33.3	90	11	4	76

II. 분석대상 및 데이터처리

➤ 데이터처리(Data Preprocessing)

(1) 데이터결합: 주은행선 좌석정보 데이터 + 시종착역별 운행정보 데이터 + 변수일원화

(2) 변수생성: 수송수요 예측에 도움될 수 있는 다양한 파생변수 후보 생성 및 32종 최종 선택

카테고리	정의
과거승차인원수	과거 1개월~12개월 전 승차인원수 데이터를 변수로 생성
시간정보	양력/음력 기반 공휴일과 대체휴일로 주말수/주중수/공휴일수 등 생성
좌석 및 운행정보 (비공식)	공급차량수, 공급좌석합계수, 승차수입금액, 승차인원수, 승차연인거리, 좌석거리, 1인당수입율, 공급대비승차율, 운행대비고객이동, 관광, 일반, 일반/관광, 대수송, 임시, 확정, 시발역, 종착역, 시발종착역, 열차운행횟수 전처리
외부환경 (코로나 이슈 제외)	코로나 관련 정부기관 대응정도와 감염/격리/사망자 수 포함한 경제환경 데이터 결합

- 과거 수송수요가 미래의 수요에 영향을 줄 수 있기 때문에 과거 1개월~12개월(1년) 전의 수송수요 생성
- 시간정보를 포함하는 시계열 데이터에서 각 월의 실제 날짜 수, 주말/주중 수, 공휴일/명절 수 등 생성
- 제공받은 수송수요를 포함한 공급좌석 정보, 열차 정보, 운행 정보 등을 재계산하여 생성
- COVID-19 종료 이벤트 반영 위해 예방접종 인원수, 격리자 수, 사망자 수, 정부대응 지수 등 결합
- 경제상황의 변동을 모델에 반영하기 위해서 한국의 주식 시장 지표와 소비자의 물가 수준 등의 지표 결합
- 변수 내 패턴이 선형이거나 스펙에 맞지 않는 이상치 변수 삭제

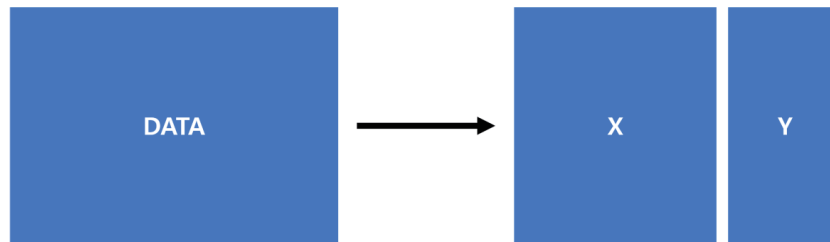
➔ 15가지 예측 케이스에 32개 변수를 모두 사용하지 않고 가장 예측 성능이 높은 조합을 선택적 활용

II. 분석대상 및 데이터처리

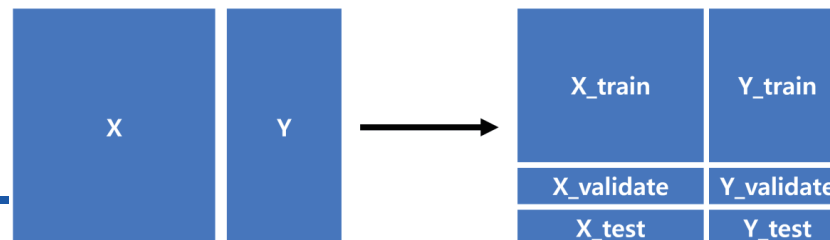
➤ 데이터처리(Data Preprocessing)

- (1) 데이터결합: 주운행선 좌석정보 데이터 + 시종착역별 운행정보 데이터 + 변수일원화
- (2) 변수생성: 수송수요 예측에 도움될 수 있는 다양한 파생변수 후보 32종을 준비
- (3) 시계열처리: 변수 별 일별데이터 집계 → 월별데이터 변환
- (4) 데이터 분할: 예측 모델링 사용을 위해 종속/독립변수와 학습(Training)/검증(Validate)/예측(Test)으로 분리

✓ 데이터 분할: (1) 목표/종속변수 Y와 설명/독립변수 X설정



✓ 데이터 분할: (2) 학습데이터와 검증 및 예측 데이터로 분할



Kyungwon Kim

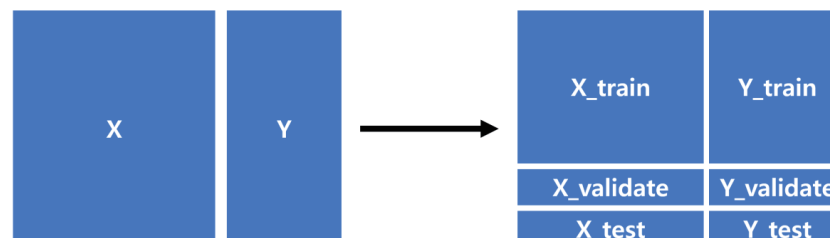
12 / 30

II. 분석대상 및 데이터처리

➤ 데이터처리(Data Preprocessing)

- (1) 데이터결합: 주운행선 좌석정보 데이터 + 시종착역별 운행정보 데이터 + 변수일원화
- (2) 변수생성: 수송수요 예측에 도움될 수 있는 다양한 파생변수 후보 32종을 준비
- (3) 시계열처리: 변수 별 일별데이터 집계 → 월별데이터 변환
- (4) 데이터 분할: 예측 모델링 사용을 위해 종속/독립변수와 학습(Training)/검증(Validate)/예측(Test)으로 분리

✓ 데이터 분할: (2) 학습데이터와 검증 및 예측 데이터로 분할



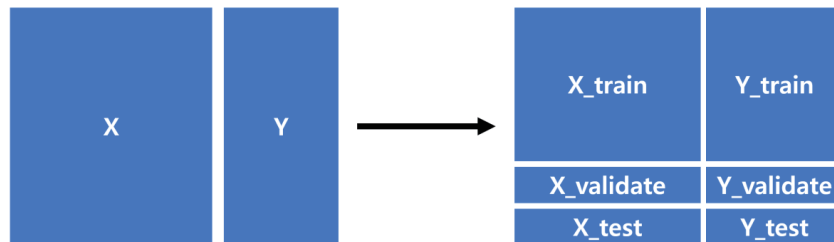
데이터 종류	기간	의미
종속변수(Y)	-	예측 대상인 승차인원수 변수
독립변수(X)	-	승차인원수에 영향을 줄거라 예상되는 변수 32종
학습(Training)	2015년 1월 ~ 2023년 3월	모델 구축을 위해 사용되는 데이터
검증(Validate)	2023년 4월 ~ 2024년 3월	모델 구축 후 예측성능 검증에 사용되는 데이터
예측(Test)	2024년 4월 ~ 2025년 12월	검증성능이 높은 모델로 실제 예측에 사용되는 데이터

II. 분석대상 및 데이터처리

➤ 데이터처리(Data Preprocessing)

- (1) 데이터결합: 주운행선 좌석정보 데이터 + 시종착역별 운행정보 데이터 + 변수일원화
- (2) 변수생성: 수송수요 예측에 도움될 수 있는 다양한 파생변수 후보 32종을 준비
- (3) 시계열처리: 변수 별 일별데이터 집계 → 월별데이터 변환
- (4) 데이터 분할: 예측 모델링 사용을 위해 종속/독립변수와 학습(Training)/검증(Validate)/예측(Test)으로 분리

✓데이터 분할: (2) 학습데이터와 검증 및 예측 데이터로 분할



데이터 종류	기간	의미
종속변수(Y)	-	예측 대상인 승차인원수 변수
독립변수(X)	-	승차인원수에 영향을 줄거라 예상되는 변수 32종
학습(Training)	2015년 1월 ~ 2023년 3월	모델 구축을 위해 사용되는 데이터
검증(Validate)	2023년 4월 ~ 2024년 3월	모델 구축 후 예측성능 검증에 사용되는 데이터
예측(Test)	2024년 4월 ~ 2025년 12월	검증성능이 높은 모델로 실제 예측에 사용되는 데이터

→ 예측 시점의 독립변수 값은, 미래 KTX 공급계획을 사용하되 각각의 미래시점 예측치를 사용

CONTENTS

I. 연구배경 및 목적

II. 분석 대상 및 데이터처리

- 분석대상(Target)
- 데이터이해(Descriptive Statistics & Visualization)
- 데이터처리(Data Preprocessing)

III. 인공지능 이해 및 예측 방법론

- 모델링 진화와 작동방식의 이해
- 머신러닝 기반 알고리즘
- 딥러닝 기반 알고리즘

IV. 성능검증 및 예측결과

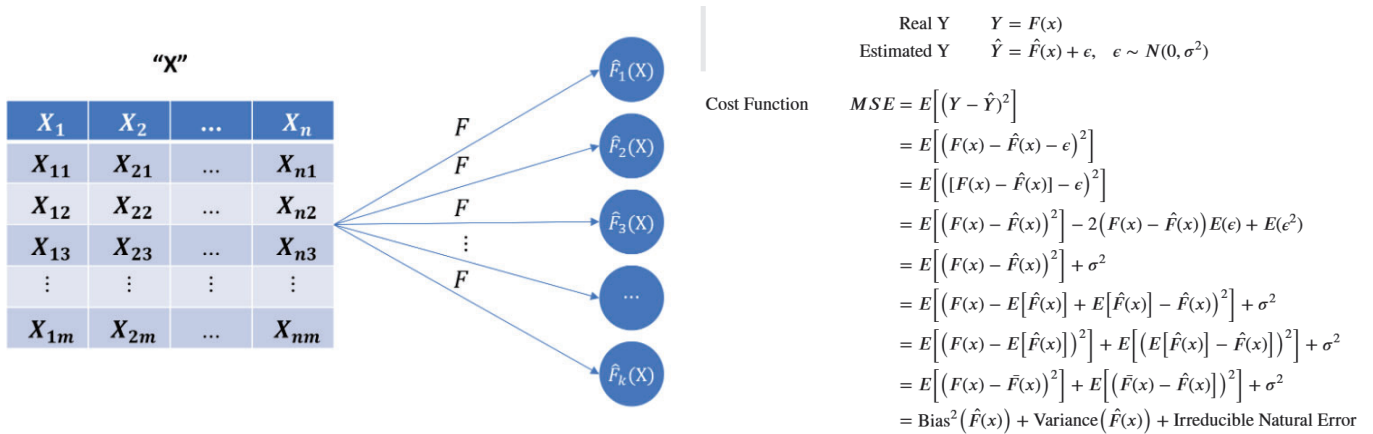
- 2024년도 노선별 예측검증 방향
- 2024년도 노선별 예측검증 요약
- 2024년도 노선별 예측검증 및 2025년도 예측
- 과거년도 대비 2025년도 예측치 일평균 비교

V. 결론 및 향후 방향

III. 인공지능 이해 및 예측 방법론

➤ 머신러닝 기반 알고리즘

- **Machine Learning 4종:** Random Forest (2001), XGBoost (2014), LGBM (2016), CatBoost (2017)
- 딥러닝 알고리즘보다 일반적으로 성능이 낮아 성능비교를 위해 활용되는 기초알고리즘(Base Algorithm)



→ ML은 실제 회귀분석의 오류를 더욱 줄이기 위해 **Bias or Variance**를 줄이는데 특화 개발된 알고리즘들

III. 인공지능 이해 및 예측 방법론

➤ 딥러닝 기반 알고리즘

- **MLP 계열 5종:** MLP (2010), NBEATS (2019), NHITS (2021), NBEATSx (2021), TiDE (2023)
- **RNN 계열 5종:** RNN (2014), LSTM (2014), GRU (2014), DilatedRNN (2017), TCN (2018)
- 데이터 패턴의 비선형 조합을 통해 높은 수준의 패턴학습을 시도하는 머신러닝 알고리즘의 발전된 버전
- 최근까지 개발된 알고리즘 중 KTX 수요예측 시계열 데이터에 적합한 알고리즘 선정

구분	CNN (Convolutional Networks)	RNN (with Attention) (Recurrent Neural Network)	MLP (Multi-Layer Perceptron)
설명	<ul style="list-style-type: none"> 합성곱 계층을 통해 데이터의 특징 추출 후, 그 특징들을 기반으로 분류하는 딥러닝 분석모델 	<ul style="list-style-type: none"> 데이터 순서정보를 반영하는 재귀 구조의 딥러닝 분석모델 Attention: 특정 데이터에 주목해 모델의 성능을 향상시키는 기법 	<ul style="list-style-type: none"> 퍼셉트론(인간의 신경계 모방)을 여러 계층으로 조합한 분석모델 딥러닝 알고리즘의 출발점
장점	<ul style="list-style-type: none"> 타 알고리즘 대비 빠른 학습속도 데이터 분할 분석을 통한 오버피팅(과적합) 가능성 감소 텍스트 분류 시, 다수의 단어 조합에 대한 패턴 고려 가능 	<ul style="list-style-type: none"> 시계열 정보(텍스트 등) 반영 가능 입-출력 개수에 따른 다양한 모델 구성 가능 (1-1, 1-N, N-1, N-N) Attention: 학습 결과에 대한 주요 활용 데이터 시각화 가능 	<ul style="list-style-type: none"> 타 분석모델 대비 간단한 구조 타 분석모델들과의 조합을 통해, 다양한 분야에서 활용 가능
단점	<ul style="list-style-type: none"> 하이퍼 파라미터 설정에 따라, 과도한 데이터 손실에 따른 언더피팅 발생 가능 	<ul style="list-style-type: none"> 입력 데이터 전처리 필수 (텍스트 임베딩 등) Attention: 컴퓨팅 자원 추가소모 	<ul style="list-style-type: none"> 모델의 복잡도(계층 수 등) 증가 시, 오버피팅 발생 가능성 급증
주요 적용분야	<ul style="list-style-type: none"> 텍스트 분류 영상인식(탁월한 성능 발휘) 	<ul style="list-style-type: none"> 사진 설명, 텍스트 분류, 번역 등 	<ul style="list-style-type: none"> 범주형 데이터 분류 / 예측

III. 인공지능 이해 및 예측 방법론

➤ 머신러닝 및 딥러닝 기반 알고리즘

모델링 종류	알고리즘 분야	이름	개발연도
Machine Learning	Bagging	RandomForest	2001
		XGBoost	2014
	Boosting	LightGBM	2016
		CatBoost	2017
Deep Learning	MLP	MLP	2010
		NBEATS	2019
		NHITS	2021
		NBEATSx	2021
		TiDE	2023
		DeepNPTS	2023
	RNN	RNN	2014
		LSTM	2014
		GRU	2014
		DilatedRNN	2017
		TCN	2018
		DeepAR	2020
	Generative AI	Vanilla Transfor	2021
		Informer	2021
		TFT	2021
		Autoformer	2021
		PatchTST	2022

➤ 예측 검증

- 6가지 검증지표: MSE, RMSE, MSPE, MAE, MAPE, MedAE, MedAPE
- **MSPE, MAPE, MedAPE** 3가지의 검증 지표는 퍼센트 기반 오류로 설명하기에 이해가 용이

CONTENTS

I. 연구배경 및 목적

II. 분석 대상 및 데이터처리

- 분석대상(Target)
- 데이터이해(Descriptive Statistics & Visualization)
- 데이터처리(Data Preprocessing)

III. 인공지능 이해 및 예측 방법론

- 모델링 진화와 작동방식의 이해
- 머신러닝 기반 알고리즘
- 딥러닝 기반 알고리즘

IV. 성능검증 및 예측결과

- 2024년도 노선별 예측검증 방향
- 2024년도 노선별 예측검증 요약
- 2024년도 노선별 예측검증 및 2025년도 예측
- 과거년도 대비 2025년도 예측치 일평균 비교

V. 결론 및 향후 방향

IV. 성능검증 및 예측결과

➤ 2024년도 노선별 예측검증 및 2025년도 예측 과정

- (1) 14종의 AI 알고리즘으로 검증기간에 예측 후 검증 지표 6종으로 성능 평가 후
- (2) 가장 성능이 높은 알고리즘으로 미래기간의 수송수요 예측

데이터 종류	기간	의미
종속변수(Y)	-	예측 대상인 승차인원수 변수
독립변수(X)	-	승차인원수에 영향을 줄거라 예상되는 변수 32종
학습(Training)	2015년 1월 ~ 2023년 3월	모델 구축을 위해 사용되는 데이터
검증(Validate)	2023년 4월 ~ 2024년 3월	모델 구축 후 예측성능 검증에 사용되는 데이터
예측(Test)	2024년 4월 ~ 2025년 12월	검증성능이 높은 모델로 실제 예측에 사용되는 데이터

IV. 성능검증 및 예측결과

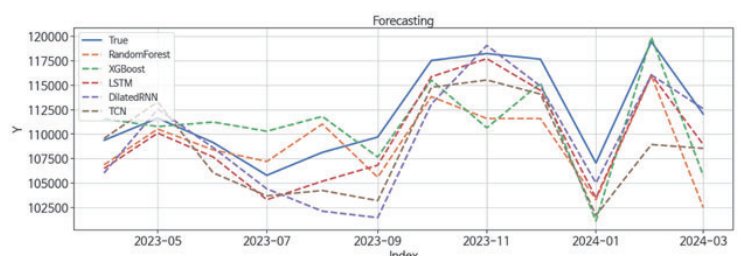
➤ 2024년도 노선별 예측검증 및 2025년도 예측 과정 (경부선 전체)

- (1) 14종의 AI 알고리즘으로 검증기간에 예측 후 검증 지표 6종으로 성능 평가 후
 - 경부선 전체의 경우 모든 알고리즘 중 NBEATS 알고리즘이 가장 높은 검증성능(적은 오차) 보임

	MSE	RMSE	MSPE	MAE	MAPE	MedAE	MedAPE
NBEATS	2.76.E+10	2.E+05	0.22%	1.E+05	3.98%	1.E+05	3.96%
NBEATSx	2.76.E+10	2.E+05	0.22%	1.E+05	3.98%	1.E+05	3.96%
NHITS	8.67.E+10	3.E+05	0.70%	3.E+05	7.50%	3.E+05	7.54%
GRU	1.24.E+11	4.E+05	0.98%	3.E+05	8.73%	3.E+05	8.44%
DilatedRNN	1.61.E+11	4.E+05	1.29%	4.E+05	10.48%	3.E+05	9.09%
RNN	1.88.E+11	4.E+05	1.54%	4.E+05	10.44%	3.E+05	10.06%
LSTM	1.21.E+12	1.E+06	10.13%	8.E+05	23.14%	4.E+05	12.54%
TCN	2.08.E+12	1.E+06	17.13%	1.E+06	28.12%	3.E+05	9.69%
TIDE	6.16.E+11	8.E+05	5.29%	7.E+05	19.62%	7.E+05	20.23%
MLP	5.93.E+11	8.E+05	4.93%	7.E+05	20.64%	8.E+05	22.93%

- 검증기간 동안 약 330만명 전후의 실제 수송수요를 NBEATS 알고리즘도 잘 예측

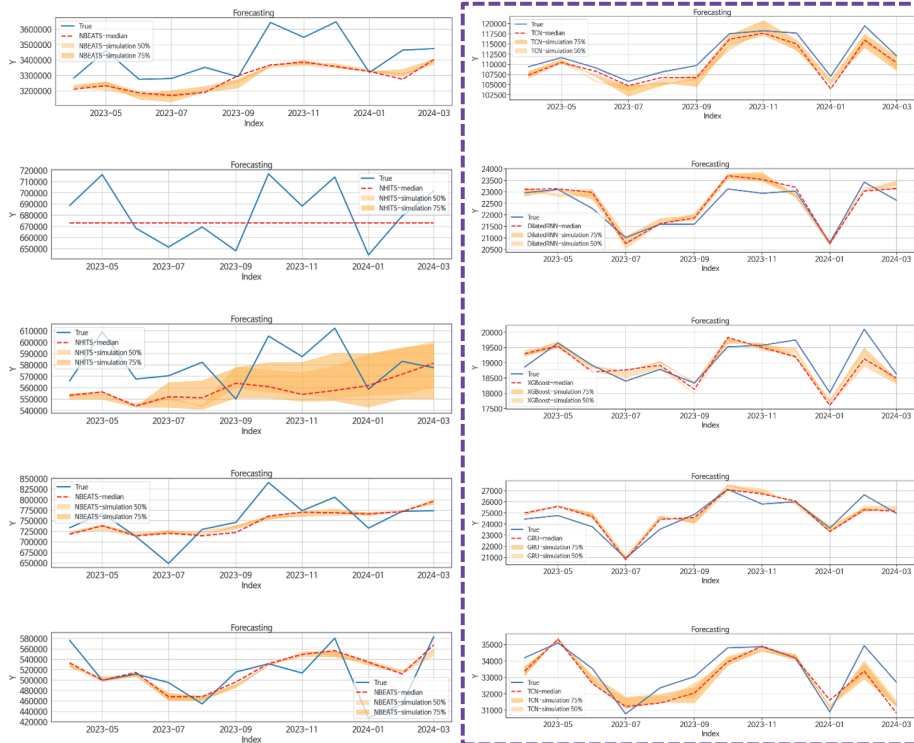
Time	TRUE	NBEATS	NBEATSx	NHITS	GRU	DilatedRNN
2023-04-01	3,281,372	3,209,957	3,209,957	3,231,787	3,094,675	3,086,898
2023-05-01	3,460,800	3,232,946	3,232,946	3,156,379	3,133,825	3,049,322
2023-06-01	3,274,361	3,186,084	3,186,084	3,061,863	3,163,710	3,036,746
2023-07-01	3,279,768	3,169,345	3,169,345	3,148,811	3,169,294	3,080,458
2023-08-01	3,352,224	3,189,312	3,189,312	3,064,115	3,056,019	3,126,249
2023-09-01	3,291,100	3,298,763	3,298,763	2,994,854	3,144,784	3,098,163
2023-10-01	3,643,417	3,367,579	3,367,579	3,108,759	2,982,193	3,062,360
2023-11-01	3,547,176	3,385,911	3,385,911	3,155,472	3,027,069	2,973,285
2023-12-01	3,647,548	3,357,167	3,357,167	3,207,876	3,108,191	2,955,161
2024-01-01	3,318,669	3,326,396	3,326,396	3,209,645	3,173,108	3,061,901
2024-02-01	3,464,111	3,274,700	3,274,700	3,292,522	3,130,876	3,102,309
2024-03-01	3,473,501	3,403,702	3,403,702	3,279,372	3,194,160	3,031,659



IV. 성능검증 및 예측결과

➤ 2024년도 노선별 예측검증 및 2025년도 예측 과정 (경부선 전체)

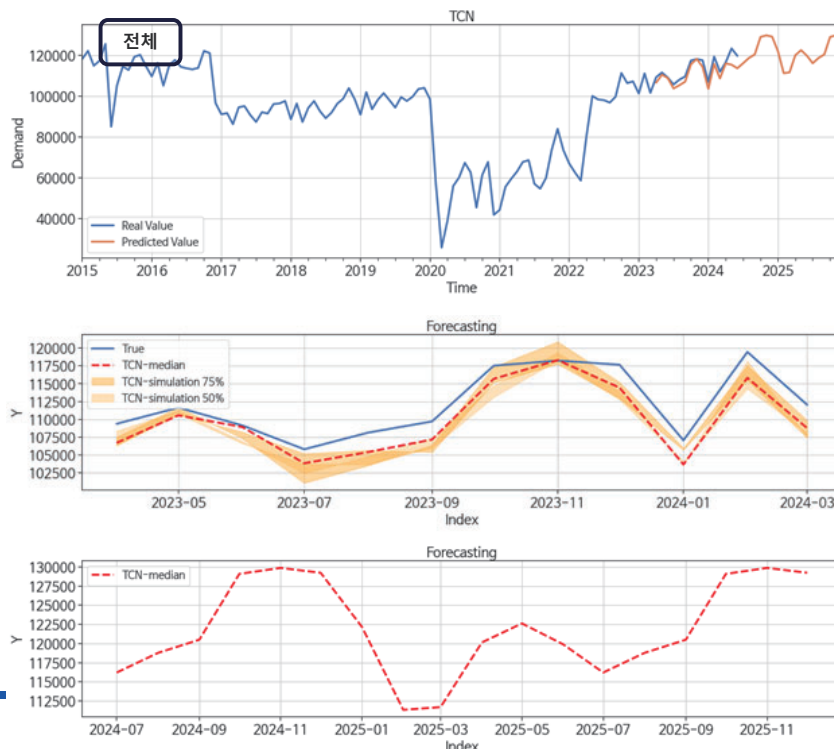
→ 예측 시점의 독립변수 값은, 미래 KTX 공급계획을 사용하되 각각의 미래시점 예측치를 사용



IV. 성능검증 및 예측결과

➤ 2024년도 노선별 예측검증 및 2025년도 예측 과정 (경부선 전체)

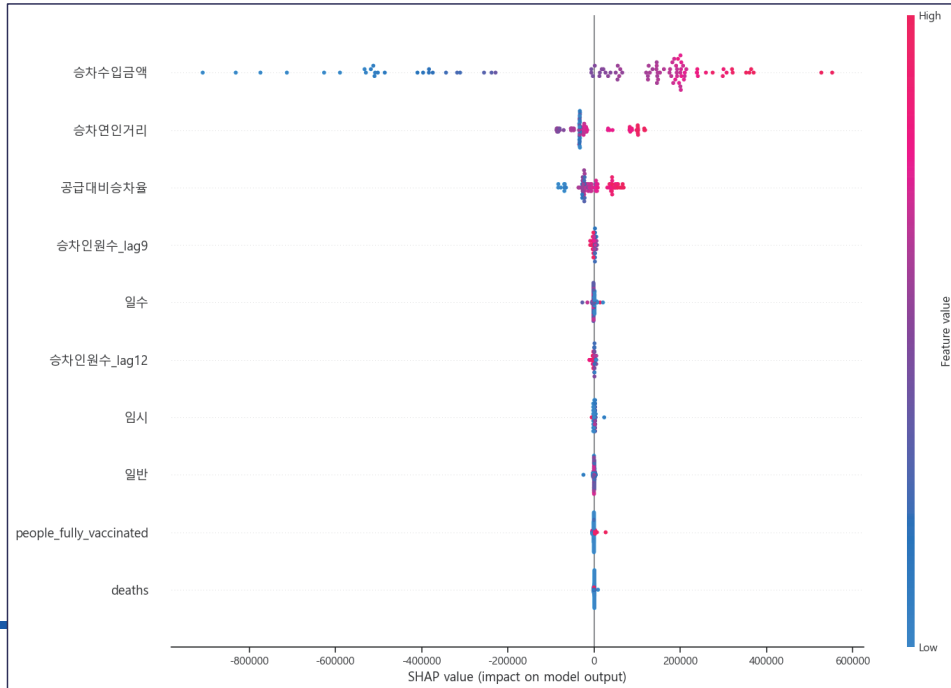
- (1) 14종의 AI 알고리즘으로 검증기간에 예측 후 검증 지표 6종으로 성능 평가 후
- (2) 가장 성능이 높은 알고리즘으로 미래기간의 수송수요 예측



IV. 성능검증 및 예측결과

➤ 2024년도 노선별 예측검증 및 2025년도 예측 과정 (경부선 전체)

- (1) 14종의 AI 알고리즘으로 검증기간에 예측 후 검증 지표 6종으로 성능 평가 후
- (2) 가장 성능이 높은 알고리즘으로 미래기간의 수송수요 예측
- (3) 필요에 따라 SHAP 알고리즘을 사용하여 변수 기여방향 확인 (전체 <<< 주말, 주중)



Kyungwon Kim

24 / 30

IV. 성능검증 및 예측결과

➤ 2024년도 노선별 예측검증 요약

- 15개 케이스 모두에서 **최고 예측 성능의 알고리즘은 딥러닝 알고리즘**
- 최고 예측성능 알고리즘은 **MLP 계열이 9개(60%)를 차지하고 RNN 계열이 6개(40%)를 차지**
- MLP 계열 성능은 **NBEATS > NHITS** 순서 + RNN 계열 성능은 **GRU = LSTM = DilatedRNN** 순서
- 노선별 예측오차는 **최소 2.35% ~ 최대 6.56% 분포며 평균적으로 4.06% 오차**
- 요일별 예측오차는 **주중(5.59%) < 주말(3.99%) < 전체(2.61%)** 순으로 **예측 성능이 높음**

주운행선	전체주말주중	과거데이터활용	미래데이터생성	알고리즘	알고리즘 계열	MSPE	MAPE	MedAPE	Average
경부선	전체	과거승차인원수	-	NBEATS	MLP	0.22%	3.98%	3.96%	2.72%
경부선	주말	좌석맞운행정보+외부환경	시간정보	DilatedRNN	RNN	0.62%	5.90%	2.84%	3.12%
경부선	주중	좌석맞운행정보	시간정보+외부환경	GRU	RNN	0.96%	8.59%	8.09%	5.88%
경전선	전체	좌석맞운행정보+외부환경	시간정보	NHITS	MLP	0.15%	3.35%	3.58%	2.36%
경전선	주말	좌석맞운행정보+외부환경	시간정보	LSTM	RNN	0.67%	6.18%	4.09%	3.65%
경전선	주중	좌석맞운행정보+외부환경+과거승차인원수	시간정보	DilatedRNN	RNN	0.71%	6.87%	6.40%	4.66%
동해선	전체	과거승차인원수	-	NHITS	MLP	0.26%	4.28%	3.70%	2.75%
동해선	주말	과거승차인원수	-	NHITS	MLP	1.15%	9.52%	9.01%	6.56%
동해선	주중	과거승차인원수	-	NBEATS	MLP	1.13%	8.61%	7.72%	5.82%
전라선	전체	과거승차인원수	-	NBEATS	MLP	0.25%	3.73%	3.07%	2.35%
전라선	주말	좌석맞운행정보+외부환경	시간정보	LSTM	RNN	0.34%	5.13%	4.84%	3.44%
전라선	주중	좌석맞운행정보+외부환경	시간정보	NBEATS	MLP	1.44%	9.51%	7.04%	6.00%
호남선	전체	과거승차인원수	-	NBEATS	MLP	0.30%	4.60%	3.74%	2.88%
호남선	주말	좌석맞운행정보+외부환경	시간정보	GRU	RNN	0.69%	5.78%	3.07%	3.18%
호남선	주중	좌석맞운행정보	시간정보+외부환경	NBEATS	MLP	1.09%	8.72%	6.98%	5.60%

Kyungwon Kim

INCHEON NATIONAL UNIVERSITY

25 / 30



Thank You

