**논문번호:** J1\_202500001

**원논문 제목:** “2025년도 KTX 수요 예측 및 정책적 의사결정 : XAI 기반 실증적 예측연구(Demand Forecast and Policy Decisions for KTX in 2025: An Empirical Forecasting Study based on XAI)”

**안녕하십니까 한국빅데이터학회 학회지 에디터님,**

원고의 재제출과 심사자의 의견에 대해 답변할 수 있는 기회를 주셔서 감사합니다.

요청해주신 바와 같이 아래에 심사자별 코멘트에 대한 답변을 포함하였습니다.

심사자분들께서 지적해 주신 내용 외에도 일부 업데이트 사항을 반영 및 표시하여 논문을 개선하였습니다.

감사합니다,

차명주, 오영택, 이승연, 김경원

## 심사자 #1

### 심사자 #1, 심사내용 #1

|  |
| --- |
| **1. 정확한 의미 전달을 위해 영어 약어(KTX, TGV, HSR 등)는 첫 언급 시 전체 명칭(예: Korea Train Express)을 함께 표기해 주시기 바랍니다.** |

**저자답변:** 영어 약어 표현을 신경 쓰지 못했는데 검토를 해주셔서 감사합니다. 모든 처음 등장하는 영어 약어를 전체 명칭을 반영하였으며 두번째로 등장하는 영어 약어에는 전체 명칭을 삭제하였습니다.

**수정사항:**

* KTX(Korea Train eXpress)는 프랑스의 TGV(Train à Grande Vitesse), 일본의 신칸센(Shinkansen), 중국의 HSR(High-Speed Rail)
* ARIMA(Autoregressive Integrated Moving Average)
* LSTM(Long Short-Term Memory)과 XGBoost(eXtreme Gradient Boosting)
* (eXplainable Artificial Intelligence, XAI)
* LIME(Local Interpretable Model-agnostic Explanations)
* Bagging(Bootstrap aggregating)
* LightGBM(Light Gradient Boosting Machine), 그리고 CatBoost(Categorical Boosting)
* LSTM
* 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)

### 심사자 #1, 심사내용 #2

|  |
| --- |
| **2. <표 2>에서는 각 수치마다 ‘%’ 기호를 반복 표기하기보다는, 열 제목 또는 표 하단에 단위를 일괄 표기하는 방식이 가독성에 더 효과적입니다. 또한 <그림 3>과 유사한 형식의 시각화로 제공할 것을 권장합니다.** |

**저자답변:** 심사내용의 업데이트를 통해 표2 뿐만 아니라 표3과 표4도 가독성을 높일 수 있었으며, Thank you for your helpful suggestion. We also apologize for the lack of readability of our answer to a similar comment in the previous review, wherein Reviewers 1 and 2 already pointed out this issue. We all agreed with the comment and thought that it was crucial. We then added mathematical formulas as possible in the first revision as follows:

**수정사항:**

|  |  |
| --- | --- |
|  |  |

### 심사자 #1, 심사내용 #3

|  |
| --- |
| **3. <그림 3(나)>, <그림 4> 등 다수의 그래프에서 x축과 y축 라벨이 단순히 'index', 'y'로 표기되어 있습니다. 보다 직관적인 변수명으로 수정하여 정보 전달력을 높여주시기 바랍니다.** |

**저자답변:** Thank you for pointing out that the algorithm’s parameter tuning and fusion effects can help our prediction. We used a general function to confirm the effect of the influencing feature (as you mentioned its significance in Concern #8) of “segment interest“ rather than the algorithm comparison. However, we felt the need to apply and reflect on these references you provided, so we added these things to the Conclusion and Future Work section as follows.

**수정사항:** We updated the manuscript as follows.

* **Add the following sentences and references in [V. CONCLUSION AND FUTURE WORK]:**
* **Before:** The development of this approach could be a potential future research.
* **After:** The ensemble or hybrid effect of algorithms can also improve prediction performance [43], [44]. Therefore, the development of these approaches has the potential for further research.

## 심사자 #2

### 심사자 #2, 심사내용 #1

|  |
| --- |
| - <그림 5>의 Feature Value 라벨은 시각성 확보를 위해 글꼴 크기를 확대해 주시기 바랍니다. |

**저자답변:** Thank you for suggesting other metrics to evaluate our proposed model. We used an additional evaluation method, Logloss, to overcome the disadvantage of the AUC score. Given that both metrics were not related to the threshold and suitable for classification with imbalanced class problems, such as CTR prediction, we selected them. We provided additional answers in Concern #2 of Reviewer #3.

### 심사자 #2, 심사내용 #2

|  |
| --- |
| - <그림 3>의 y축 단위는 ‘4x10⁶’과 같이 지수 형태로 표기하면 수치의 범위와 해석에 있어 더 명확하게 전달됩니다. |

**저자답변:** As in Concern #1, we agree that the performance should be evaluated using various metrics. Thank you for pointing out such a critical issue. We provided additional answers below following our previous ones.

### 심사자 #2, 심사내용 #3

|  |
| --- |
| - <표 3>과 같은 주요 표는 독자의 이해를 돕기 위해 시각화(예: 막대그래프)로 병행 제시하는 것이 효과적입니다. 아래는 예시 코드이며, 실제 데이터에 맞게 수정해 사용해 주시기 바랍니다. import pandas as pd import matplotlib.pyplot as plt plt.rcParams['font.family'] = 'NanumGothic'  data = { "노선": ["경부선", "경전선", "동해선", "전라선", "호남선"], "알고리즘": ["LightGBM", "LightGBM", "XGBoost", "LSTM", "XGBoost"], "MSPE(%)": [0.30, 0.94, 0.91, 0.99, 0.37], "MAPE(%)": [4.39, 9.35, 9.18, 9.77, 5.22], "MedAPE(%)": [2.79, 9.22, 8.40, 9.92, 5.17], "평균오차(%)": [2.49, 6.50, 6.16, 6.89, 3.59] }  df = pd.DataFrame(data)  # 그래프 생성 plt.figure(figsize=(10, 6)) bar\_width = 0.2 index = range(len(df))  plt.bar(index, df["MSPE(%)"], bar\_width, label="MSPE") plt.bar([i + bar\_width for i in index], df["MAPE(%)"], bar\_width, label="MAPE") plt.bar([i + bar\_width\*2 for i in index], df["MedAPE(%)"], bar\_width, label="MedAPE") plt.bar([i + bar\_width\*3 for i in index], df["평균오차(%)"], bar\_width, label="평균오차")  plt.xlabel("노선") plt.ylabel("오차율 (%)") plt.title("KTX 노선별 예측 성능 비교") plt.xticks([i + bar\_width\*1.5 for i in index], df["노선"]) plt.legend() plt.tight\_layout() plt.grid(axis='y', linestyle='--', alpha=0.7)  plt.show() |

**저자답변:** As in Concern #1, we agree that the performance should be evaluated using various metrics. Thank you for pointing out such a critical issue. We provided additional answers below following our previous ones.

### 심사자 #2, 심사내용 #4

|  |
| --- |
| - <표 4>과 같은 주요 표는 독자의 이해를 돕기 위해 시각화(예: 막대그래프)로 병행 제시하는 것이 효과적입니다. 아래는 예시 코드이며, 실제 데이터에 맞게 수정해 사용해 주시기 바랍니다. import matplotlib.pyplot as plt plt.rcParams['font.family'] = 'NanumGothic'  labels = ["경부선", "경전선", "동해선"] year\_2019 = [3000064, 552302, 470230] year\_2024 = [3418760, 675125, 573080] year\_2025 = [3261319, 624792, 521400]  x = range(len(labels)) bar\_width = 0.25  # 막대그래프 그리기 plt.figure(figsize=(10, 6)) plt.bar(x, year\_2019, width=bar\_width, label='2019년') plt.bar([i + bar\_width for i in x], year\_2024, width=bar\_width, label='2024년') plt.bar([i + bar\_width \* 2 for i in x], year\_2025, width=bar\_width, label='2025년 예측')  plt.xlabel("노선") plt.ylabel("승차 인원수 (명)") plt.title("KTX 노선별 연도별 승차 인원수 비교") plt.xticks([i + bar\_width for i in x], labels) plt.legend() plt.grid(axis='y', linestyle='--', alpha=0.7)  plt.tight\_layout() plt.show() |

**저자답변:** As in Concern #1, we agree that the performance should be evaluated using various metrics. Thank you for pointing out such a critical issue. We provided additional answers below following our previous ones.

⬛