**논문번호:** J1\_202500002

**원논문 제목:** “설명가능한 인공지능 활용 서울시 지하철 혼잡도 예측과 정책적 의사결정 지원을 위한 비즈니스 애널리틱스(Business Analytics to Predict Seoul Subway Congestion and Policy Decision-making using Explainable AI)”

**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,

박재흥, 김경원

## 심사자 #1

### 심사자 #1, 심사내용 #1

|  |
| --- |
| : 선행연구와의 차별성을 표 형태 등으로 정리하여 **본 연구의 기여 지점**을 보다 명확히 도식화할 것. |

**저자답변:** 본 연구의 기여점을 강조할 수 있도록 중요한 코멘트를 주셔서 감사합니다. 기존의 지하철 혼잡도 예측이 다소 전통적이며 제한적인 알고리즘을 사용하고 있었고 설명가능한 인공지능을 활용한 해석에 관련된 연구는 전무한 상황이었습니다. 그러한 부분을 강조하기 위해서 아래와 같이 표를 반영하고 관련 내용들을 본문에 추가함으로써 독자들의 이해도를 높일 수 있었습니다. 감사합니다.

**추가사항:**

|  |
| --- |
| 또한 지하철 뿐만 아니라 버스나 교통상황들을 예측하기 위해 도메인 특성에 맞는 딥러닝 알고리즘들을 개발하여 높은 성능을 달성한 연구들도 많아지고 있다 [20-21]. 그 외에 관련 선행연구는 <표 1>의 연구목적 별로 상세하게 확인할 수 있다. 전반적으로 머신러닝과 딥러닝 알고리즘을 활용한 혼잡도 예측 연구가 성능이 증가되고 있음을 알 수 있다.  인공지능의 대표적 알고리즘인 머신러닝과 딥러닝은 높은 성능의 예측값을 제공하지만, 예측의 근거를 사람이 직관적으로 이해하기 어렵다는 한계가 있다. 특히 머신러닝과 딥러닝으로 진화하면서 “블랙박스(Black Box)” 모델로 불릴 정도로 내부 의사결정 과정이 불투명해, 사용자가 모델의 결정 이유를 해석하거나 검증하기 어렵다. 이러한 한계를 극복하고자 최근 설명 가능한 인공지능(Explainable AI, XAI) 에 대한 관심이 높아지고 있다. XAI는 머신러닝, 딥러닝 등 복잡한 인공지능 알고리즘의 결과를 사람이 이해할 수 있는 형태로 설명하는 기술이다 [27]. 단순히 “무엇을 예측했는가”를 넘어 “왜 그렇게 예측했는가”를 설명할 수 있게 하며, 모델 개발자 뿐만 아니라 최종 사용자, 정책 입안자, 서비스 관리자에게 중요한 정보를 제공한다. 이를 통해 사용자는 모델이 어떤 특성과 패턴을 기반으로 결정을 내렸는지 알 수 있으며 모델의 신뢰성(Trustworthiness)과 투명성(Transparency)를 높이는데 기여한다 [28]. 지하철 혼잡도와 관련되서는 아직까지 XAI를 활용한 연구들이 전무하지만, 다른 교통 분야에서는 인공지능 기반의 높은 성능 뿐만 아니라 해석까지 포함하고 있는 연구들이 늘어나고 있다. 특히 버스의 승객수를 예측하거나 [22], 인구밀도 [23]에 활용될 뿐만 아니라 교통상황과 사고를 상세하게 이해하기 위한 노력들이 증가하고 있다 [24-26].  하지만 XAI를 활용하기 위해선 분석 모델링 이후 별도의 추가 계산이 필요하고 방법론에 따라서는 전체(Global) 설명보다는 부분적인 설명에 그치며 설명 결과가 비전문가에게는 다소 어려울 수 있다. 또한 모든 데이터와 모델에 완벽히 적용되지는 않으며, 복잡한 상호작용에 대한 설명과 실무 적용에는 여전히 기술적 도전이 요구되는 상황이다. 그럼에도 불구하고 이러한 연구들이 확산되고 여러 연구자들에게 이해도를 높일 수 있는 연구들이 많이 필요하다. 대표적인 XAI 기법에는 반사실적 설명(Counterfactual Explanations), 변수 중요도(Feature Importance), LIME (Local Interpretable Model-agnostic Explanations), SHAP (Shapley Additive Explanation) 등이 있다. 본 연구에서는 이러한 방법들 중 가장 기술적으로 완성도가 높은 SHAP (Shapley Additive explanation) 방법을 이용하여 고성능의 지하철 혼잡도 예측의 근거를 확인하고 변수들의 기여도를 분석한다. 게임 이론 기반의 세플리 값(Shapley Value)을 활용해 각 변수의 기여도를 계산하므로 직관적이고 일관된 설명을 제공한다. 지하철 혼잡도 예측을 위해 기존에는 정성적인 접근 또는 제한된 인공지능 알고리즘을 활용하였지만, 본 연구는 대표적인 머신러닝과 딥러닝을 모두 활용하는 예측 방법론을 제공하는 동시에 설명가능한 인공지능을 통해 지하철 혼잡도 대응을 위한 접근을 추가함으로써 기존 문헌들을 확장하였다. |

### 심사자 #1, 심사내용 #2

|  |
| --- |
| **②**: '설명 가능한 인공지능(XAI)' 이론 및 SHAP 알고리즘에 대한 설명이 단편적임. 설명 가능성의 **정의, 장단점, 한계**등을 보완적으로 정리할 것. |

**저자답변:** 중요한 지적을 해 주셔서 감사합니다. 익숙하지 않은 독자들에게 이해도를 높일 수 있도록 관련 배경과 정의, 그리고 장점과 단점 및 한계점 등을 보완하여 본 연구의 목적과 알고리즘 선택 기준 등을 상세화 할 수 있었습니다. 감사합니다.

**수정사항:**

|  |  |
| --- | --- |
| **[기존]**  인공지능의 대표적 알고리즘인 머신러닝과 딥러닝은 높은 성능의 예측값을 제공하지만, 예측 근거를 사람이 이해하기 어렵다는 한계가 있다. 블랙박스와 같은 이러한 한계를 극복하고자 설명 가능한 인공지능 (Explainable AI, XAI)의 필요성이 높아지고 있다. 설명 가능한 인공지능은 복잡한 인공지능 알고리즘의 결과를 사람이 이해할 수 있도록 설명하는 방법으로 [9], 사용자는 모델이 어떤 특성과 패턴을 기반으로 결정을 내렸는지 알 수 있어서 신뢰성을 높일 수 있다 [10]. 본 연구에서는 설명 가능한 인공지능 방법 중 대표적인 SHAP (Shapley Additive explanation) 방법을 이용하여 고성능의 지하철 혼잡도 예측의 근거를 확인하고 변수들의 기여도를 분석한다. | **[변경]**  인공지능의 대표적 알고리즘인 머신러닝과 딥러닝은 높은 성능의 예측값을 제공하지만, 예측의 근거를 사람이 직관적으로 이해하기 어렵다는 한계가 있다. 특히 머신러닝과 딥러닝으로 진화하면서 “블랙박스(Black Box)” 모델로 불릴 정도로 내부 의사결정 과정이 불투명해, 사용자가 모델의 결정 이유를 해석하거나 검증하기 어렵다. 이러한 한계를 극복하고자 최근 설명 가능한 인공지능(Explainable AI, XAI) 에 대한 관심이 높아지고 있다. XAI는 머신러닝, 딥러닝 등 복잡한 인공지능 알고리즘의 결과를 사람이 이해할 수 있는 형태로 설명하는 기술이다 [27]. 단순히 “무엇을 예측했는가”를 넘어 “왜 그렇게 예측했는가”를 설명할 수 있게 하며, 모델 개발자 뿐만 아니라 최종 사용자, 정책 입안자, 서비스 관리자에게 중요한 정보를 제공한다. 이를 통해 사용자는 모델이 어떤 특성과 패턴을 기반으로 결정을 내렸는지 알 수 있으며 모델의 신뢰성(Trustworthiness)과 투명성(Transparency)를 높이는데 기여한다 [28]. 지하철 혼잡도와 관련되서는 아직까지 XAI를 활용한 연구들이 전무하지만, 다른 교통 분야에서는 인공지능 기반의 높은 성능 뿐만 아니라 해석까지 포함하고 있는 연구들이 늘어나고 있다. 특히 버스의 승객수를 예측하거나 [22], 인구밀도 [23]에 활용될 뿐만 아니라 교통상황과 사고를 상세하게 이해하기 위한 노력들이 증가하고 있다 [24-26].  하지만 XAI를 활용하기 위해선 분석 모델링 이후 별도의 추가 계산이 필요하고 방법론에 따라서는 전체(Global) 설명보다는 부분적인 설명에 그치며 설명 결과가 비전문가에게는 다소 어려울 수 있다. 또한 모든 데이터와 모델에 완벽히 적용되지는 않으며, 복잡한 상호작용에 대한 설명과 실무 적용에는 여전히 기술적 도전이 요구되는 상황이다. 그럼에도 불구하고 이러한 연구들이 확산되고 여러 연구자들에게 이해도를 높일 수 있는 연구들이 많이 필요하다. 대표적인 XAI 기법에는 반사실적 설명(Counterfactual Explanations), 변수 중요도(Feature Importance), LIME (Local Interpretable Model-agnostic Explanations), SHAP (Shapley Additive Explanation) 등이 있다. 본 연구에서는 이러한 방법들 중 가장 기술적으로 완성도가 높은 SHAP (Shapley Additive explanation) 방법을 이용하여 고성능의 지하철 혼잡도 예측의 근거를 확인하고 변수들의 기여도를 분석한다. 게임 이론 기반의 세플리 값(Shapley Value)을 활용해 각 변수의 기여도를 계산하므로 직관적이고 일관된 설명을 제공한다. 지하철 혼잡도 예측을 위해 기존에는 정성적인 접근 또는 제한된 인공지능 알고리즘을 활용하였지만, 본 연구는 대표적인 머신러닝과 딥러닝을 모두 활용하는 예측 방법론을 제공는 동시에 설명가능한 인공지능을 통해 지하철 혼잡도 대응을 위한 접근을 추가함으로써 기존 문헌들을 확장하였다. |

### 심사자 #1, 심사내용 #3

|  |
| --- |
| **③**: 변수 생성 및 병합의 기준, 결측치 처리 기준, 이상치 처리 방식 등에 대해 **명확한 기술적 설명**을 추가할 것. |

**저자답변:** 본 연구의 주요 목적인 공공 융합 데이터베이스를 활용한 지하철 혼잡도 예측을 위해 데이터들의 병합 과정을 자세하게 기술하지 못하여 독자들이 향후 동일한 실험을 진행하기 어려울 것이라는 중요한 포인트를 지적해 주셔서 정말 감사합니다. 따라서 저희 연구진은 기존에 작성된 본문의 내용을 전면 수정하여 병합 과정이 Concat인지 Merge인지 어떠한 방법론을 사용했는지, Merge의 경우 어떠한 Key를 사용했는지, 그리고 데이터베이스 로딩 과정에서의 불일치성을 어떻게 동일하게 변경했는지 등의 관점들을 반영하여 전처리 진행 과정을 아래와 같이 전면 수정하였습니다. 감사합니다.

**수정사항:**

|  |  |
| --- | --- |
| **[기존]**  본 연구에서 “서울교통공사\_1\_8호선 역별 일별 시간대별 승객유형별 승하차인원” 데이터베이스를 기준데이터로 사용하였다. 해당 데이터는 교통카드를 이용한 지하철 호선별 역별(서울시 관할 1~9호선) 시간대별 승하차인원을 나타내는 정보로 수송일자, 호선명, 역명, 승하차구분, 승객유형, 시간대별 승하차인원을 알 수 있다. 여기서 변수를 가공하여 시간대별 승하차인원의 합과 차, 시간대별 우대권인원수, 시간대별 청소년인원수 변수를 추가하였다. “서울교통공사\_지하철혼잡도정보” 데이터베이스를 5시부터 23시까지 시간대별 상선 혼잡도와 하선 혼잡도로 분리하여 위 기준 데이터에 병합하였다. 그 다음으로 "서울교통공사\_월별 환승유입인원" 데이터베이스의 역 및 월별 환승유입인원 변수를 기준 데이터에 병합하였다. 그 후 2024년 6월 30일 기준 "서울교통공사\_역사운영 현황" 데이터베이스에서 역별 면적, 출입구, 섬식여부, 환승노선 갯수를 추출해서 기준 데이터에 병합하였다. 그리고 종속변수로 사용된 승강장혼잡도를 면적 변수와 승하차인원(합) 변수를 사용하여 생성하였다. 마지막으로 날짜의 경우, 위 데이터에서 수송일자를 활용하여 년도, 월, 일, 요일, 주 변수를 만들었고, 파이썬 holidayskr 라이브러리를 통해 공휴일 변수를 생성할 수 있었다. | **[변경]**  본 연구에서 서울교통공사에서 제공하는 “서울교통공사\_1\_8호선 역별 일별 시간대별 승객유형별 승하차인원” 분기별 데이터베이스 3년치를 수집하여 기준데이터로 사용하였다. 해당 데이터는 날짜, 호선, 역번호, 역명, 승차여부, 하차여부, 승객유형에 따른 시간대별 승하차인원으로 구성되어 있다. 분기별로 변수명이 상이한 경우 동일하게 변경하였고, 특정 분기 데이터에서 24시간대 이후 변수가 존재하지만 약 46만개의 결측치로 구성되어 있어 제거하였다. 여기에서 승차인원수, 하차인원수 변수를 생성하고 동일 시간대의 승차 및 하차인원수를 합/차를 계산하여 승하차인원합/승하차인원차 변수를 추출하였다. 그리고 승객유형 정보에서 우대권을 사용한 승하차 인원과 청소년 승하차 인원을 추출하여 변수로 반영하였다. 다음으로 서울교통공사가 제공하는 “서울교통공사\_지하철혼잡도정보” 데이터베이스는 30분 단위로 측정 후 연도별로 제공하는데, 일부 시간대에서 36~1133개 정도의 결측치가 확인되었다. 전체 샘플에 비해서는 미비하기에 해당 결측치는 제거하였고, 날짜, 호선, 역명, 상행여부, 하행여부에 따른 시간대별 혼잡도 지표로 구성되어 있다. 여기에서 시간대별 상선 혼잡도와 하선 혼잡도를 추출하여 변수로 구성하였다. 그런 다음 “날짜”를 공통 키(Key) 값으로 사용하여 기준데이터에 병합(Left Join)하였다. 다음으로 "서울교통공사\_월별 환승유입인원" 데이터베이스는 연도별로 제공하는데, 2021년도의 데이터에서 최대 4개의 결측치를 발견하였고 나머지 연도에서는 결측치가 존재하지 않았다. 마찬가지로 결측치의 개수가 매우 미비하여 제거하였다. 그리고 2021년도는 월별 환승유입인원이 열로 반영되어 있어서 행으로 구성되어 있는 2022년과 2023년도와 구조를 맞추기 위해 행으로 변환후 병합(Concat)하였다. 최종적으로 날짜, 호선, 역명, 환승유입인원수로 구성되어 있고 “날짜”를 공통 키(Key)로 활용하여 앞선 데이터베이스들과 병합(Left Join)하였다. 그리고 2024년 6월 30일 기준 "서울교통공사\_역사운영 현황" 데이터베이스에서 호선, 역명, 면적, 층수, 승강장유형, 출입구, 환승노선을 알 수 있다. 여기에서 결측치는 166개였는데 이는 환승노선이 없는 역을 의미하기에 0으로 처리하였다. 따라서 날짜별 역별 면적, 층수, 섬식여부와 같은 승강장유형, 출입구, 환승노선 갯수를 추출할 수 있었으며 “날짜”를 공통 키(Key) 값으로 사용하여 앞선 데이터베이드글에 병합(Left Join)하였다. 그리고 기존에 추출한 승하차인원합 변수에서 면적을 나누어서 종속변수로 생성할 수 있었다. 마지막으로 모든 데이터베이스의 공통 킷값으로 사용된 “날짜”를 활용하여 년도, 월, 일, 요일, 주 변수를 만들었고, 파이썬 holidayskr 라이브러리를 통해 공휴일 변수를 생성하여 시계열 정보들을 지하철 혼잡도 예측에 활용할 수 있었다. |

### 심사자 #1, 심사내용 #4

|  |
| --- |
| **④**: ‘승강장 혼잡도 = 면적 대비 승하차 인원’이라는 정의에 대한 **이론적 또는 정책적 타당성**을 보강할 것. |

**저자답변:** 분석 대상의 이론적 근거를 탄탄하게 하여 독자들이 연구의 목적을 잘 이해할 수 있는 검토의견을주셔서 감사합니다. 본 연구에서 면적 대비 승하차인원을 종속변수로 고려하게 된 선행연구와 고민들을 포함한 내용들을 본문에 추가하였습니다. 감사합니다.

**추가사항:**

|  |
| --- |
| 교통상태의 혼잡을 완화하는 연구는 지속적으로 해결하려는 난제이다. 교통 시설이나 수단, 그리고 이용대상에 따라서 관련된 이론이나 방법론들이 다양하게 개발되고 있다. 이호 외(2015)은 승강장 혼잡도를 승강장 면적 대비 승차대기인원으로 정의하였다. 이 때 승차대기인원은 열차 탑승을 위해 대기하는 인원으로 열차 출발정보와 승강장 누적진입인원을 활용하였다 [5]. 이상준 외(2021)은 승강장 내 승객수와 승객점유면적의 비율을 사용하여 혼잡도를 정의하였다. 승강장 내 승객이 실제 이용하는 대기 면적 개념을 제안하였으며, 승차승객 대기면적계수를 추정하였다. 승차승객 대기면적계수의 경우, 상대식 승강장은 0.291, 섬식 승강장은 0.270을 적용하여 혼잡도를 산정하였다 [6]. 이상준 외(2020)은 1분 단위 승객 통행량과 승강장 통행량을 산출하고 승강장별 실용대기면적을 나누어 승강장 동적 혼잡도를 추정하였다 [7].  관련 연구들을 종합하면, 지하철 혼잡도는 크게 세 가지로 구분될 수 있다. 첫 번째, 승강장 내에서 지하철을 승하차 하기 위해 대기할 때의 혼잡도이다. 둘째는 지하철에 탑승한 후 객차 내에서 승객들이 겪는 혼잡도이다. 공공데이터포털의 “서울교통공사\_지하철혼잡도정보” 데이터베이스가 수집하는 혼잡도는 정원 대비 승차인원으로 정의하고 승차인과 좌석수가 일치할 경우 혼잡도 34%로 산정한다. 세번째는 앞서 설명한 2가지의 혼잡도를 모두 혼합한 것이다. 지하철 전체 객차에 승객이 대기할 때의 혼잡도와 승하차가 이루어지는 일련의 과정에서 느끼는 객차 내 혼잡도를 모두 포함하여 승강장 내에서 겪는 승객들의 모든 혼잡도이다. 단순히 객차 안으로 한정할 경우 객차의 공간이 모두 점유가 되었다고 하더라도 객차 외부에서 대기하는 승객들의 피로도를 반영하지 못할 수 있다. 반면 객차 외부로만 한정할 경우 사회적 이슈로 객차가 혼잡하지 않고 외부 승강장 내 혼잡이 발생하는 이슈를 반영하지 못한다. |

### 심사자 #1, 심사내용 #5

|  |
| --- |
| **⑤**: 각 알고리즘별 하이퍼파라미터 튜닝 전략, 교차검증 여부, 조기 종료 등 **과적합 방지 기법**을 명확히 기술할 것. |

**저자답변:** 일반적으로 회귀문제에서 Train만 성능이 높고 Test가 성능이 좋지 않은 과적합을 방지하는 것은 것은 중요한 문제입니다. 하지만 본 연구의 실험 결과 중 <표 4>를 보면 Test 대비 Train만 과도하게 성능이 좋다라고 판단하기 어렵습니다. Metric에 따라 다르지만 MSPE와 MAPE는 Train과 Test의 성능 차이가 상위 알고리즘에선 5% 정도의 차이밖에 되지 않습니다. MedAPE의 경우엔 Random Forest를 제외한 모든 알고리즘의 Test 성능이 Train보다 오히려 높게 나타나기도 합니다. 이러한 이유로 인해 특별히 과적합을 방지하기 위한 실험을 추가하진 않았습니다. 또한 본 연구의 목적은 여러 개의 DB를 활용한 예측 성능의 향상 그리고 설명가능한 인공지능을 활용하여 의사결정을 지원하는 것이 더욱 주된 목적이기 때문에 알고리즘의 하이퍼파라미터는 누구나 사용할 수 있는 default 값을 사용했습니다. 단, 딥러닝 알고리즘에서 dropout ratio와 조기 종료 등을 적용하긴 했습니다. 하지만 알고리즘의 예측 성능을 극대화하기 보단 여러가지 DB 사용을 통한 예측 성능 향상과 XAI 기반 비즈니스 애널리틱스가 주된 목적이기에 과적합에 대한 고려를 하지 않았음을 이해해 주시기 바랍니다.

### 심사자 #1, 심사내용 #6

|  |
| --- |
| **⑥**: 시계열적 데이터임에도 불구하고 시점 간 상관 구조를 고려하지 않은 점은 한계임. **시계열 특성 반영 여부**를 언급하거나, 한계점으로 명시할 것. |

**저자답변:** 본 연구에서 활용하는 시간 단위가 연도, 월, 주별이 아닌 “시간”단위로 예측합니다. 그리고 지하철 혼잡도의 도메인 특성상 직전 시점들의 영향보다 해당 시간대나 시간에서의 환경적 요인에 영향을 많이 받기 때문에, 직전 시점들의 영향을 고려하는 직접적인 시계열 알고리즘을 활용하기 보다 해당 시간대나 시간의 정보를 전처리 과정에서 반영후 비시계열 알고리즘으로 예측을 하였습니다. 실제로 LSTM, GRU와 같은 시계열 알고리즘의 사용이 직전 시점 데이터에 영향을 받을 가능성이 높은 월별이나 연도별 예측과 같은 수요예측에서는 성과가 있었지만 각 시간대별 지하철 혼잡도 예측 성능에서는 효과가 없었습니다. 그러한 이유로 전처리 과정에서 시간별로 Train과 Test를 구분하지 않고 랜덤하게 데이터를 분리 후 시간 패턴 전처리 데이터를 반영하여 해당 시간대의 환경 데이터 기반으로 예측함으로써 예측 성능을 향상시킬 수 있었습니다.

**수정사항:**

|  |  |
| --- | --- |
| **[기존]**  마지막으로 평균적으로 역마다 약 5.49만개 정도의 샘플들을 8:2의 비율로 구분하여 훈련 (Training) 세트와 테스트 (Test) 세트로 분리하였다. | **[변경]**  마지막으로 평균적으로 역마다 약 5.49만개 정도의 샘플들을 랜덤하게 8:2의 비율로 구분하여 훈련 (Training) 세트와 테스트 (Test) 세트로 분리하였다. 연구에 활용한 데이터는 시계열이지만 실제 혼잡도를 예측하는 시간 단위는 월별이나 주별이 아닌 “시간”이다. 그리고 지하철의 특성상 바로 직전 시간의 정보가 현재와 미래 시간의 혼잡도에 영향을 주기보다 출근시간대, 퇴근시간대, 휴일 등과 같이 특정 상황에 따라 혼잡도가 더 영향을 줄 가능성이 높다. 또한 각 시간별 과거 3년치의 패턴이 반영되는 만큼 유사한 시간이나 시간대별 샘플들이 풍부하다. 따라서 바로 직전 시점들의 영향을 고려한 시계열 알고리즘을 직접적으로 사용하기 보다 시계열의 패턴을 파생변수로 반영하는 것이 적합할 것으로 생각된다. 그러므로 시간적인 순서로 데이터를 분리하지 않고 랜덤하게 데이터를 분리하였고, 시간 정보를 반영하기 위해 후술될 전처리 과정에서 연도, 월, 요일, 주, 공휴일 여부 등의 시간패턴을 반영한다. |

### 심사자 #1, 심사내용 #7

|  |
| --- |
| **⑦**: SHAP 분석 결과 해석 시, 변수 기여의 **방향성과 중요도**에 대한 **일관된 해석 기준**을 제시할 것. |

**저자답변:** 독자들이 SHAP 결과의 해석과 의사결정을 위해 중요한 지적을 해 주셔서 감사합니다. 익숙하지 않은 독자들에게 이해도를 높일 수 있도록 <표 6>을 통해 일관된 해석을 할 수 있는 방법과 향후 활용을 정리하여 표현하였으며, 본문에도 표에 대한 설명 그리고 각 Decision Plot과 Summary Plot 기반 해석을 하기 전에도 내용을 추가하여 목적과 해석을 더욱 쉽게 할 수 있도록 내용을 추가했습니다. 감사합니다.

**추가사항:**

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| SHAP 알고리즘은 개별 지하철 혼잡도에 대한 설명을 Decision Plot으로 제공해주며, 이를 누적하여 전체 지하철 혼잡도는 Summary Plot으로 표현해준다. Decision Plot은 X축은 해당 변수에 대한 혼잡도 예측에 미치는 정도인 SHAP 값을 표현하며, Y축은 변수의 중요도를 내림차순으로 표현한다. 따라서 표시된 선그래프는 제일 아래에 있는 변수의 예측 기여 SHAP 값을 위로 누적시키며 최종적인 혼잡도 예측을 표현한다. 실제로 위로 올라갈수록 이전 SHAP 값 대비 기울기가 커지듯 그만큼 혼잡도 예측에 영향이 크다고 판단할 수 있다. 그러므로 개별 혼잡도에서 어떤 변수들이 주요 결정 요인인지 파악하는데 용이하다. Summary Plot의 X축은 Decision Plot과 마찬가지로 해당 변수의 예측 기여 SHAP 값이다. Y축도 변수의 중요도를 내림차순으로 표현하긴 하지만 변수별로 샘플기여 SHAP 값을 누적하여 분포로 표현한다. 따라서 개별기여 값을 점으로 표현하고 색상을 통해 값의 변화에 대한 예측 기여 방향을 표현한다. 그리고 분포변화가 클수록 예측에 크게 기여하는 것이기 때문에 위로 갈수록 혼잡도 예측에 크게 기여한다. 그러므로 전체 예측에 대한 주요변수 우선순위와 해석으로 모델의 전반적인 이해와 의사결정에 활용될 수 있다. 좀 더 상세한 변수의 중요도와 방향성 그리고 해석과 활용에 대해 <표 7>에 정리하였다.  <표 7> SHAP 알고리즘이 제공하는 Decision Plot과 Summary Plot의 해석방법과 설명 그리고 활용목적 요약   |  |  |  | | --- | --- | --- | | **구분** | **Decision Plot** | **Summary Plot** | | **대상** | 개별 샘플 | 전체 데이터 | | **해석방법** | **> X축:** 변수의 예측 기여 SHAP 값  **> Y축:** 변수중요도 내림차순 기반 누적기여 SHAP 값  **> 선그래프:** 기울기 기반 해당변수 영향력  (기울기 변화가 클수록 예측에 크게 기여하기에 위로 갈수록 중요) | **> X축:** 변수의 예측 기여 SHAP 값  **> Y축:** 변수중요도 내림차순 기반 변수별 샘플기여 누적 SHAP 값  **> 분포선그래프:** 변수값의 변화에 따른 예측기여 방향과 정도  - **파란색:** 변수의 낮은값에 대한 SHAP 값 분포  - **빨간색:** 변수의 높은값에 대한 SHAP 값 분포  - **파란색->빨간색 분포변화:** 예측에 대한 긍정 기여  - **빨간색->파란색 분포변화:** 예측에 대한 부정 기여  (분포변화가 클수록 예측에 크게 기여하기에 위로 갈수록 중요변수) | | **설명** | 개별 예측의 결정 경로를 누적 SHAP 값으로 시각화  (선=개별기여, 변수=결정경로 우선순위) | 개별 예측의 변수별 기여방향을 누적하여 SHAP 값 분포로 시각화  (점=개별기여, 색=기여방향, 변수=전체결정 우선순위) | | **활용목적** | 1) 개별 예측의 주요변수 식별과 결정과정 추적  2) 샘플결과간 비교를 통한 개별샘플 인싸이트 확보 | 1) 전체 예측의 주요변수 식별과 영향방향 해석  2) 모델의 전반적인 이해로 의사결정 지원 |   <표 7>의 해석방법에 따르면 특정 샘플의 최종 혼잡도가 얼마나 예측이 되었고 어떤 이유로 예측값이 나왔는지 결정경로를 선그래프로 표현하고 있으며, 선의 기울기에 따라 기여 방향을 확인할 수 있다.  각 샘플마다의 지하철 혼잡도 예측 설명력을 누적한 요약 플롯(Summary Plot)을 통해 전체의 일반화된 설명력을 확인할수 있다. 그리고 <표 7>의 해석방법에 따르면 각 변수들마다 지하철 혼잡도에 어떠한 우선순위와 방향으로 기여를 하고 있는지 분포선그래프로 표현하고 있으며, 색상의 변화에 따르 기여 방향을 일반화 할 수 있다.  관련 설명은 <표 7>을 통해 상세히 요약되어 있다. |

### 심사자 #1, 심사내용 #8

|  |
| --- |
| **⑧**: 특정일에 대한 해석뿐만 아니라, **누적된 결과 기반의 인사이트**를 보다 정량적으로 도출할 수 있는 방식 제시 필요. |

**저자답변:** Decision Plot의 경우 기울기의 변화를 통해 기여정도를 해석할 수 있고, Summary Plot은 변수별색상변화를 통해 기여정도를 해석할 수 있습니다. 관련 설명을 <표 6>을 통해 상세히 제시하였습니다. 그런데 기울기의 변화를 정량적으로 표현할 수 있겠지만 X 또는 Y축 중 하나만 정량적인 수치로 표현되어 있어서 기울기를 수치화 하는 것이 어려운 상황입니다. 마찬가지로 변수별 색상변화 대비 SHAP 값의 분포 변화를 직선처럼 표시를 하였지만 실제로는 다차원 공간의 변화이기 때문에 정량적으로 특정 숫자로만 표현하는 것이 기준에 따라 너무 다양할 수 있고 실제를 왜곡할 수 있는 해석이라 조심해야 하는 상황입니다. 이러한 이유로 “기울기의 변화” 그리고 “변수별 색상변화”로만 해석할 수밖에 없는 점을 이해하 주시기 바랍니다. 향후 그러한 기술들이 추가될 필요성은 매우 공감하는 바입니다. 감사합니다.

### 심사자 #1, 심사내용 #9

|  |
| --- |
| **⑨**: 혼잡도 완화 방안(예: 환승노선 확충, 섬식 플랫폼 확대)에 대해 **SHAP 기반 변수 해석과의 정합성**을 논리적으로 연결할 것. |

**저자답변:** 결론의 해석이 실험 결과와 어떻게 연결되는지 독자들이 쉽게 확인하기 어려울 수 있음을 발견했습니다. 따라서 실험 결과를 요약하고 어떻게 해석을 하게 되는지 그 과정을 추가하였으며 좀 더 의사결정을 지원하기 위해 어떠한 기획이 가능할 수 있는지 방향을 제시함으로써 SHAP 기반 변수 해석을 더욱 충분히 할 수 있었습니다. 감사합니다.

**수정사항:**

|  |  |
| --- | --- |
| **[기존]**  혼잡도를 낮추기 위해서는 승하차 인원을 줄이는 것이 근본적인 해결책일 수 있지만, 그 외에도 환승노선을 늘리는 사업이 확충되거나 섬식형태가 아닌 승강장 플랫폼을 확장하는 것이 방법일 수 있다. 또한, 한주의 출근이 몰리는 월요일과 화요일에 혼잡도가 높아지기 때문에 근무의 유연화 문화를 통해 혼잡도의 개선이 가능할 것으로 생각된다. 본 연구는 설명 가능한 인공지능을 활용하여 지하철 혼잡도를 실시간으로 빠르게 확인할 수 있는데 도움을 줄 수 있으므로, 서비스 측면에서 시민들에게 빠르고 효과적인 대응책을 마련하는 데 도움이 될 것이다. 대표적 딥러닝 알고리즘의 성능이 높게 나타나고 있어 향후 더욱 정교한 알고리즘으로 성능을 향상히켜 볼 계획이며 추가적으로 시계열 예측으로 확장한다면 실제 미래 전망에도 활용도가 높을것으로 예상된다. | **[변경]**  본 연구의 XAI 실험 결과에 따르면 “하차인원 > 승차인원 > 출입구 > 섬식여부 > 월요일 > 화요일”의 순서로 혼잡도를 높이며, “환승노선\_개수 > 환승유입인원수 > 하선/상선 혼잡도 > 주 > 년도 > 월 > 일요일 > 목요일 > 혼잡도”의 순서 변수들이 혼잡도를 낮추는 요인들로 분석되었다. 따라서 혼잡도를 낮추기 위해서는 하차/승차 인원을 줄이는 것이 가장 근본적인 해결책일 수 있지만, 그 외에도 지하철 노선을 확충하거나 노선간 연결점들을 늘려 환승노선과 환승유입인원수를 늘리는 방법도 활용할 수 있다. 지하철의 출입구를 줄이거나 섬식이 아닌 승강장 플랫폼 형태로 지하철을 개발하는 것이 필요한데, 이것은 환경적 비용적으로 고려가 되어야 하는 부분이라 기존 지하철 환경의 변경 보단 신규 노선들에 반영되면 좋을 수 있는 인싸이트로 볼 수 있다. 또한 주의 출근이 몰리는 월요일과 화요일에 특히 혼잡도가 높아지기 때문에 근무의 유연화 문화를 확산하여 혼잡도 개선이 가능할 것이다. 그리고 연초보다 연말로 갈수록 혼잡도가 낮아지는 경향이 있기 때문에 시민들을 대상으로 하는 행사들도 하반기로 배치하는 것이 혼잡도를 낮추는데 기여할 수 있다. 그리고 과거보다 현재로 갈수록 지하철 혼잡도는 낮아지는 추세가 있다. 이는 여러가지 종류의 교통 개발과 정책적 접근이 효과를 가져오는 것으로 판단할 수 있는 근거자료가 될 수 있다. 이처럼 본 연구는 설명 가능한 인공지능을 활용하여 지하철 혼잡도를 실시간으로 빠르게 확인할 수 있는데 도움을 줄 수 있으므로, 서비스 측면에서 시민들에게 빠르고 효과적인 대응책을 마련하는 데 도움이 될 것이다. 대표적 딥러닝 알고리즘의 성능이 높게 나타나고 있어 향후 더욱 정교한 알고리즘으로 성능을 향상히켜 볼 계획이며 추가적으로 시계열 예측으로 확장한다면 실제 미래 전망에도 활용도가 높을것으로 예상된다. |

### 심사자 #1, 심사내용 #10

|  |
| --- |
| **⑩**: ‘근무 유연화’ 제안은 유의미하나, 해당 변수와 혼잡도의 **실증적 상관관계 분석**이 포함되어야 함. |

**저자답변:** 본문에 지하철 혼잡도 예측을 위한 변수명과 예시는 설명이 되어 있으나 데이터의 특성을 이해하기 위한 통계량이 없어서 독자들이 데이터를 이해하기 어려울 수 있음을 발견했습니다. 감사합니다. 따라서 사용한 종속변수와 독립변수의 상관성을 확인하기 위해 아래와 같이 상관관계 수치를 직관적으로 이해할 수 있는 시각화를 추가하였고 관련 설명을 본문에도 추가하였습니다. 감사합니다.

**추가사항:**

|  |
| --- |
| 그리고 승강장혼잡도와 나머지 독립변수들과의 관계성을 확인하기 위해 상관관계 분석 시각화를 <그림 1>에 제시하였다.  **<그림 1> 혼잡도 예측에 사용한 변수들의 상관관계** |

### 심사자 #1, 심사내용 #11

|  |
| --- |
| **⑪: 도식, 표, 그림의 캡션 및 본문 연결 표현을 학술지 수준에 맞게 정리할 것.** |

**저자답변:** 몇가지 그림표기와 본문의 연결 방식이 학술지 기준대로 작성되지 않음을 발견하여 모두 수정하였습니다. 감사합니다.

### 심사자 #1, 심사내용 #12

|  |
| --- |
| **⑫**: 전체 문장 중 구어적 표현(예: "이러한 점에서 의미가 있다")을 **학술적 문체**로 수정할 것. |

**저자답변:** 학술적 문제는 주장을 설득력 있게 전달하는데 중요함에도 일부 구어적 표현이 있음을 발견했습니다. 이러한 부분들을 아래와 같이 모두 수정하였습니다. 감사합니다.

**수정사항:**

|  |  |
| --- | --- |
| **[기존]**  향후 수도권 인근 신도시의 개발이나 다양한 형태의 교통수단이 개발될 예정이지만 혼잡도 증가가 해소될 수 있을지 예상하기 어렵고 모든 시간대의 혼잡도를 줄이기는 쉽지 않을 것으로 생각된다.  이러한 사회적 이슈들을 개선하기 위해서는 다수가 이용하는 대표적 교통수단인 지하철의 혼잡도를 사전에 예측하는 것은 서비스의 효율성과 그리고 시민들의 편의성과 안정성 측면에서 중요한 의미가 있다  본 연구를 위해 각 데이터베이스에서 추출할 수 있는 변수들을 생성하였고 필요에 따라서는 여러 개의 데이터베이스 변수들을 결합하여 생성하기도 하였다.  그리고 셰플리 값은 협력 게임 이론 (Coalitional Game Theory)을 기반으로 변수들의 기여도를 계산하는 지표로 생각할 수 있다.  그 외 나머지 변수들은 혼잡도에 크게 기여하지 못하는 모습이다.  사회적 이슈들을 개선하기 위해서는 다수가 이용하는 대표적 교통수단인 지하철의 혼잡도를 사전에 예측하는 것은 서비스의 효율성과 그리고 시민들의 편의성과 안정성 측면에서 중요한 의미가 있다. | **[변경]**  향후 수도권 인근 신도시의 개발이나 다양한 형태의 교통수단이 개발될 예정이지만 혼잡도 증가가 해소될 수 있을지 예상하기 어렵고 모든 시간대의 혼잡도를 줄이기는 용이하지 않을 것으로 판단된다.  이러한 사회적 이슈들을 개선하기 위해서는 다수가 이용하는 대표적 교통수단인 지하철의 혼잡도를 사전에 예측하는 것은 서비스의 효율성과 그리고 시민들의 편의성과 안정성 측면에서 중요한 의의가 있다  본 연구를 위해 각 데이터베이스에서 추출할 수 있는 변수들을 생성하였고 필요에 따라서는 여러 개의 데이터베이스 변수들을 결합하여 생성하였다.  그리고 셰플리 값은 협력 게임 이론 (Coalitional Game Theory)을 기반으로 변수들의 기여도를 계산하는 지표다.  그 외 나머지 변수들은 혼잡도에 크게 기여하지 못한다.  사회적 이슈들을 개선하기 위해서는 다수가 이용하는 대표적 교통수단인 지하철의 혼잡도를 사전에 예측하는 것은 서비스의 효율성과 그리고 시민들의 편의성과 안정성 측면에서 중요한 의의가 있다. |

## 심사자 #2

### 심사자 #2, 심사내용 #1

|  |
| --- |
| 선행연구가 절대적으로 부족합니다. 혼잡도 관련된 연구들을 보완하시고, 데이터 방법론 관련된 부분, 정책적 시사점 관련된 논문들을 최소 20편이상을 추가해 주십시오. 기존 연구와의 차별성을 충분히 논의할 필요가 있음.  특히 해외 관련 사례들 연구도 추가하여 우리나라 정책들과 비교분석도 결론에서 추진할 필요가 있습니다.  1) 지하철 혼잡도 예측 관련 선행연구  2) 머신러닝 및 딥러닝 기법을 활용한 교통 혼잡도 예측 연구  3) 인공지능(XAI)의 교통 분야 적용 연구 |

**저자답변:** 결과의 가독성을 높이는데 의견 주셔서 정말 감사합니다. 그림 5의 모든 Y축의 폰트를 증가시켰습니다. Figure 공간 상의 제약으로 인해 폰트가 증가되면 라벨들이 겹치기도 하고 상대적으로 색상표시와 다른 내용들이 작아지는 제약과 논문 다단의 공간상 제약으로 무작정 폰트를 증가시킬 수는 없었음을 이해해 주시면 감사합니다. 추가적으로 폰트 크기 외에 선명도를 높여서 가독성을 높였습니다.

**수정사항:** 심사자님의 이해를 위해 그림 5(그림 8로 변경)의 경부선에 아래와 같이 변경하였음을 근거로 첨부합니다. Revision을 통해 그림이 3개가 추가됨으로써 그림의 넘버는 본문에는 변경되었으니 참고하시기 바랍니다.

### 심사자 #2, 심사내용 #2

|  |
| --- |
| 연구의 차별성 및 기여도를 보다 명확히 제시해 주십시오.  1. 6개의 공공데이터베이스를 융합한 접근법이 어떻게 기존 연구와 차별성  2. 95% 이상의 예측 정확성이 기존 연구들(예: Jang 외(2023)의 80% 정확도)과 비교, 이의 의의  3. SHAP를 활용한 설명 가능한 인공지능 접근법이 기존 블랙박스 모델과의 차별성 |

**저자답변:** 본 연구의 기여점을 강조할 수 있도록 중요한 코멘트를 주셔서 감사합니다. 기존의 지하철 혼잡도 예측이 다소 전통적이며 제한적인 데이터와 알고리즘을 사용하고 있었고 설명가능한 인공지능을 활용한 해석에 관련된 연구는 전무한 상황이었습니다. 그리고 기존 연구들에서 얼마나 제한적인 데이터와 알고리즘을 사용했는지 그리고 성능은 어떠했는지 등을 <표 1>에 반영함으로써 기존 연구 대비 본 연구의 기여를 독자들이 쉽게 이해할 수 있게 되었습니다. 아울러 SHAP 알고리즘의 설명을 추가하여 필요성과 차별성을 강조할 수 있었습니다. 감사합니다.

**추가사항:**

|  |
| --- |
| 또한 지하철 뿐만 아니라 버스나 교통상황들을 예측하기 위해 도메인 특성에 맞는 딥러닝 알고리즘들을 개발하여 높은 성능을 달성한 연구들도 많아지고 있다 [20-21]. 그 외에 관련 선행연구는 <표 1>의 연구목적 별로 상세하게 확인할 수 있다. 전반적으로 머신러닝과 딥러닝 알고리즘을 활용한 혼잡도 예측 연구가 성능이 증가되고 있음을 알 수 있다.  인공지능의 대표적 알고리즘인 머신러닝과 딥러닝은 높은 성능의 예측값을 제공하지만, 예측의 근거를 사람이 직관적으로 이해하기 어렵다는 한계가 있다. 특히 머신러닝과 딥러닝으로 진화하면서 “블랙박스(Black Box)” 모델로 불릴 정도로 내부 의사결정 과정이 불투명해, 사용자가 모델의 결정 이유를 해석하거나 검증하기 어렵다. 이러한 한계를 극복하고자 최근 설명 가능한 인공지능(Explainable AI, XAI) 에 대한 관심이 높아지고 있다. XAI는 머신러닝, 딥러닝 등 복잡한 인공지능 알고리즘의 결과를 사람이 이해할 수 있는 형태로 설명하는 기술이다 [27]. 단순히 “무엇을 예측했는가”를 넘어 “왜 그렇게 예측했는가”를 설명할 수 있게 하며, 모델 개발자 뿐만 아니라 최종 사용자, 정책 입안자, 서비스 관리자에게 중요한 정보를 제공한다. 이를 통해 사용자는 모델이 어떤 특성과 패턴을 기반으로 결정을 내렸는지 알 수 있으며 모델의 신뢰성(Trustworthiness)과 투명성(Transparency)를 높이는데 기여한다 [28]. 지하철 혼잡도와 관련되서는 아직까지 XAI를 활용한 연구들이 전무하지만, 다른 교통 분야에서는 인공지능 기반의 높은 성능 뿐만 아니라 해석까지 포함하고 있는 연구들이 늘어나고 있다. 특히 버스의 승객수를 예측하거나 [22], 인구밀도 [23]에 활용될 뿐만 아니라 교통상황과 사고를 상세하게 이해하기 위한 노력들이 증가하고 있다 [24-26].  하지만 XAI를 활용하기 위해선 분석 모델링 이후 별도의 추가 계산이 필요하고 방법론에 따라서는 전체(Global) 설명보다는 부분적인 설명에 그치며 설명 결과가 비전문가에게는 다소 어려울 수 있다. 또한 모든 데이터와 모델에 완벽히 적용되지는 않으며, 복잡한 상호작용에 대한 설명과 실무 적용에는 여전히 기술적 도전이 요구되는 상황이다. 그럼에도 불구하고 이러한 연구들이 확산되고 여러 연구자들에게 이해도를 높일 수 있는 연구들이 많이 필요하다. 대표적인 XAI 기법에는 반사실적 설명(Counterfactual Explanations), 변수 중요도(Feature Importance), LIME (Local Interpretable Model-agnostic Explanations), SHAP (Shapley Additive Explanation) 등이 있다. 본 연구에서는 이러한 방법들 중 가장 기술적으로 완성도가 높은 SHAP (Shapley Additive explanation) 방법을 이용하여 고성능의 지하철 혼잡도 예측의 근거를 확인하고 변수들의 기여도를 분석한다. 게임 이론 기반의 세플리 값(Shapley Value)을 활용해 각 변수의 기여도를 계산하므로 직관적이고 일관된 설명을 제공한다. 지하철 혼잡도 예측을 위해 기존에는 정성적인 접근 또는 제한된 인공지능 알고리즘을 활용하였지만, 본 연구는 대표적인 머신러닝과 딥러닝을 모두 활용하는 예측 방법론을 제공하는 동시에 설명가능한 인공지능을 통해 지하철 혼잡도 대응을 위한 접근을 추가함으로써 기존 문헌들을 확장하였다. |

### 심사자 #2, 심사내용 #3

|  |
| --- |
| 결과 해석과 정책 제언에 있어서 해외 사례와의 비교 보완이 필요함. |

**저자답변:** 표 3에 대한 성능 비교 시각화를 그림 5로써 추가함으로써 독자들의 이해도를 높일 수 있었습니다감사합니다.

**수정사항:** 그림 5가 아래와 같이 추가되었음을 근거로 첨부합니다.

|  |
| --- |
|  |

⬛