

XGBoost를 활용한 첨두, 비첨두시간 사고 심각도 영향요인 분석

Analysis of the Impact Factors of Peak and Non-peak Time Accident Severity Using XGBoost

성제민¹ · 윤병조^{2*}Je Min Seong¹, Byoung Jo Yoon^{2*}¹Researcher, College of Urban Science, Incheon National University, Incheon, Republic of Korea²Professor, College of Urban Science, Incheon National University, Incheon, Republic of Korea

*Corresponding author: Byoung Jo Yoon, bjoon63@inu.ac.kr

ABSTRACT

Purpose: The number of registered vehicles in Korea continues to increase. As traffic volume increases gradually due to improved quality of life, the severity of accidents is expected to increase and congestion problems are also expected. Therefore, it is necessary to analyze the accident factors of pointed traffic accidents and non-pointed traffic accidents. **Method:** The severity of the apical and non-pointed traffic accidents in Incheon Metropolitan City is analyzed by dividing them into apical and non-pointed traffic accidents to investigate the factors affecting the accident. XGBoost machine learning techniques were applied to analyze the severity of pointed and non-pointed traffic accidents and visualized as plot through the results. **Result:** It was analyzed that during non-peak hours, such as the case of the victim's vehicle type at peak times, the victim's vehicle type and construction machinery are variables that increase the severity of the accident. **Conclusion:** It is meaningful to derive the seriousness factors of apical and non-pointed accidents, and it is hoped that it will be used to reduce congestion costs by reducing the seriousness of accidents in the case of apical and non-pointed in the future.

Keywords: Machine Learning, XGBoost, Accident Severity, Variable Importance, Peak, Non-Peak

요약

연구목적: 국내의 차량 등록 대수는 계속 증가하고 있다. 삶의 질 향상으로 인한 교통량 또한 점진적으로 증가하므로 사고 심각도가 증가 및 혼잡문제 또한 야기될 것으로 예상된다. 따라서, 첨두 교통사고와 비첨두 교통사고의 사고 요인을 분석할 필요가 있다고 판단된다. **연구방법:** 인천광역시의 첨두 및 비첨두 교통사고의 심각도를 첨두와 비첨두로 나누어 분석하여 사고에 영향을 미치는 요인을 알아보고자 한다. XGBoost 머신러닝 기법을 적용하여 첨두 및 비첨두 교통사고 심각도를 분석하였으며 결과를 통하여 plot으로 시각화하였다. **연구결과:** 첨두시 피해운전자 차종_승합인 경우 등 비 첨두시는 피해운전자 차종_건설기계 등이 사고 심각도를 높게 되는 변수인 것으로 분석되었다. **결론:** 첨두와 비첨두 사고 심각도의 요인을 도출한 것에 의의가 있고 추후 첨두 및 비첨두시의 사고 심각도를 낮추고 국내 교통의 혼잡 요인을 분석하여 혼잡 비용을 줄일 수 있는 것에 활용되기를 바란다.

핵심용어: 머신러닝, XGBoost, 사고심각도, 변수중요도, 첨두, 비첨두

Received | 21 May, 2024

Revised | 20 June, 2024

Accepted | 20 June, 2024

OPEN ACCESS



This is an Open-Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/4.0>) which permits unrestricted noncommercial use, distribution, and reproduction in anymedium, provided the original work is properly cited.

서론

국내 자동차 누적 등록대수는 e-나라지표 통계청을 통하여 21년 12월 기준으로 2,491만대를 기록하고 있다. 16년 2,180만대, 18년 2,320만대, 20년 2,436만대로 자동차 누적 등록대수는 2년마다 대략 2.2%만큼 점진적으로 증가하는 추세를 보이고 있다. 또한, 삶의 질 향상으로 여가통행과 물류 통행의 증가로 인해 자동차 통행량도 증가하고 있는 추세이다. 또한, 인구의 대도시 집중현상이 증가하고 있기에 대도시권 및 도시권의 자동차 통행의 수요를 공급이 수용하지 못하는 현상이 발생하고 있고, 첨두시간의 반복적인 교통혼잡이 발생하고 있다. 본 연구는 국내의 첨두시간 교통혼잡 및 교통사고가 증가하고 있으므로 첨두시간 교통사고 심각도 분석을 진행하고 심각도에 영향을 미치는 요인이 무엇인지 분석하고자 한다. 교통사고 심각도 분석을 위해 XGBoost를 활용하여 인천광역시에서 발생하는 첨두시간 교통사고를 분석하여, 심각도 모형을 구축하여 심각도에 영향을 미치는 요인을 분석한다. 본 연구의 결과를 통해 첨두시간 교통사고 및 첨두시간 교통 혼잡에 대하여 개선하는 방안을 제시하고 다른 지역의 첨두시간 교통사고의 개선사업 및 기초적인 자료로 활용할 수 있는 방안을 마련하고자 한다.

첨두교통사고 및 비첨두 교통사고 데이터 통계분석

분석개요

본 연구는 인천광역시에서 발생한 첨두 교통사고와 비첨두 교통사고를 대상으로 첨두와 비첨두의 교통사고의 원인을 비교 분석하였다. 사고 심각도 분석을 위해 TASS의 통계 자료를 통해 2021년부터 2023년까지의 총 3년간의 인천광역시에서 발생한 첨두 교통사고 데이터와 비첨두 교통사고 데이터를 수집했다. 첨두 시간은 오전 7시부터 9시까지 설정하였고, 오후 시간은 17시부터 20시까지 설정하였고, 그 외 시간은 비첨두로 설정하였다. 수집 결과로 인천광역시의 발생한 첨두 교통사고 및 비첨두 교통사고 건수는 총 23,255건으로 이를 통하여 분석에 활용했다. 첨두 교통사고와 비첨두 교통사고의 심각도를 비교하기 위해서는 첨두 교통사고와 비첨두 교통사고의 비교를 하기 위해서 총 건수에서 첨두와 비첨두를 나누었다. 첨두 교통사고 건수는 10,508건이고 비첨두 교통사고 건수는 12,747건으로 데이터를 통하여 분석에 활용했다.

Table 1은 TAAS에서 제공한 데이터를 통해 년도, 사고 건수, 사망자 수, 부상자 수를 나누어 표현한 것이다. 사망자 수 및 부상자 수는 가해자, 피해자 모두를 합한 값을 활용했다.

분석결과

Table 1. Status of peak traffic accidents and non-peak traffic accidents in Incheon metropolitan city

구분	첨두시간 교통사고				비첨두시간 교통사고						
	사고 건수		사망자 수		부상자 수		사고 건수		사망자 수		부상자 수
	(건)	(명)	(%)	(명)	(%)	(건)	(명)	(%)	(명)	(%)	
2021	3,662	39	1.06	4,919	134.33	4,250	65	1.53	5,811	136.73	
2022	3,433	29	0.84	4,649	135.42	4,262	66	1.55	5,884	138.06	
2023	3,413	31	0.91	4,479	131.23	4,235	52	1.23	5,868	138.56	

Table 1의 첨두시간 교통사고 분석결과로 2021년의 첨두시간 교통사고는 사고 건수 대비 사망자 수의 비율이 1.06%이고 부상자 수는 134.33%로 분석되었다. 2022년의 첨두시간 교통사고는 사고 건수 대비 사망자 수의 비율이 0.84%이고 부상자 수는 135.42%로 분석되었으며, 2023년 첨두시간 교통사고는 사고 건수 대비 사망자 수의 비율이 0.91%이고 부상자 수는 131.23%로 분석되었다.

Table 1의 비첨두시간 교통사고 분석결과로 2021년의 비첨두시간 교통사고는 사고 건수 대비 사망자 수의 비율이 1.53%이고 부상자 수는 136.73%로 분석되었다. 2022년의 비첨두시간 교통사고는 사고 건수 대비 사망자 수의 비율이 1.55%이고 부상자 수는 138.06%로 분석되었으며, 2023년 비첨두시간 교통사고는 사고 건수 대비 사망자 수의 비율이 1.23%이고 부상자 수는 138.56%로 분석되었다. Table 1의 부상자 수의 비율을 확인해 보면 100%가 넘는 결과 값이 도출되었는데, 이는 가해 차량과 피해 차량을 구분하지 않고 통계를 내었기 때문에 이와 같은 결과값이 도출되었다.

Fig.1은 2021년부터 2023년까지 총 3년간의 교통사고 현황을 분석한 결과이다. 사고 건수는 줄어들고 있음을 보여주고 있지만 사망자 수와 부상자 수는 교통사고 한 건당 2명은 다치는 것으로 분석되었다. 첨두시간 보다 비첨두 시간의 사고 심각도가 높은 이유는 첨두시간의 교통량보다 비첨두 시간의 교통량이 적기 때문에 자동차 통행의 통행 속도 증가 및 사망자 수와 부상자 수가 높을 수 밖에 없는 것으로 판단 되었다. 사고 건수와 사망자 수, 부상자 수는 소폭 감소하는 것으로 보이고 있으나 비율은 차이가 없는 것으로 보이고 있다. 따라서, 사고 심각도는 줄지 않는 큰 문제점으로 보이고 있다.

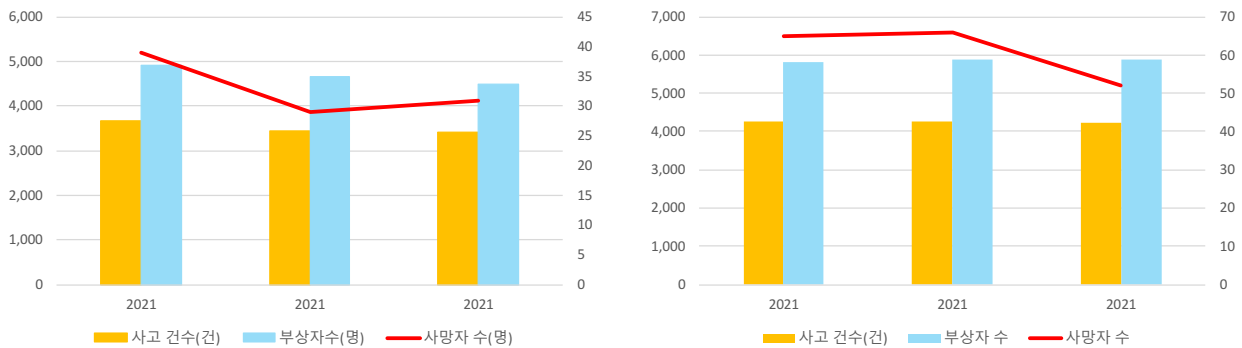


Fig. 1. Variable importance (peak), Variable importance (non-peak)

XGBOOST

XGBoost는 단일 모델을 사용하는 것 보다 복수의 모델을 활용하여 앙상블 학습을 통해 예측력을 높일 수 있는 머신러닝이다. XGBoost는 효율성과 유연성 및 휴대성이 뛰어나도록 최적화되어 다양하게 활용되고 있다. XGboost 머신러닝의 앙상블은 다른 머신러닝과는 다르게 베깅 앙상블을 사용하지 않고 부스팅을 활용하여 더 성능을 효율성있게 만들어진 모델이다. 따라서 다른 머신러닝들의 알고리즘의 단점을 보완해주기 위해 나오게 되었다. XGboost의 앙상블의 특징은 다른 머신러닝 알고리즘보다 빠르게 분석할 수 있고 과적합을 방지할 수 있도록 만들어져있다. 분류 방법과 회귀 방법을 둘다 가능하게 하기에 예측성능이 뛰어나다고 할 수 있다. 또한, 의사결정 나무 기반 앙상블 머신러닝 알고리즘이므로 회귀 모형과 분류 모형을 개발하여 우수한 성능을 보이고 변수의 중요도를 정량적으로 측정이 가능하므로 사고심각도를 예측하기에 효율적인 분석이 가능한 모델이다.

XGBoost를 이용하여 침두 교통사고 심각도와 비침두 교통사고 심각도 모형 설계 과정 결과는 아래와 같다.

본 연구는 전국 2021년부터 2023년까지 발생한 침두 교통사고와 비침두 교통사고의 심각도 모형을 구축하기 위해 XGBRegressor를 사용했다. XGBoost 또한 최적의 상태를 설정하여 모형을 이용하기 위해 하이퍼 파라미터를 Table 2와 같이 설정하였다. n_estimators는 XGBoost를 구성하는 결정 트리의 개수이다. learning_rate는 학습을 진행할 때마다 적용하는 학습률이다. subsample는 데이터 샘플링 비율 지정값이다. colsample_bytree는 트리 생성에 필요한 피쳐의 샘플링에 사용하는 것이다.

max_depth는 결정 트리의 최대 깊이를 나타낸다. gamma는 트리의 리프 노드를 추가 적으로 나눌지를 결정하는 최소 손실 감소 값이다. 침두시간의 하이퍼파라미터는 n_estimators 500, learning_rate 0.01, subsample 0.4, colsample_bytree 0.5, max_depth 4, gamma 1로 RMSE가 3.77으로 최적으로 분석되었다. 비침두시간의 하이퍼파라미터는 n_estimators 500, learning_rate 0.01, subsample 0.4, colsample_bytree 0.6, max_depth 5, gamma 0.5로 RMSE가 3.36으로 최적으로 분석되었다.

Table 2. XGBoost hyper parameters

	구분	값
과정1	n_estimators	100, 300, 500
	learning_rate	0.01, 0.5, 0.1
과정2	subsample	0.2, 0.3, 0.4
	colsample_bytree	0.5, 0.6, 0.7
과정3	max_depth	3, 4, 5
	gamma	0.5, 1, 1.5

XGBOOST 변수중요도

최적의 파라미터를 적용한 후 모형을 구축하여 XGBoost 자체 모형으로 변수중요도를 분석하였다.

모형 예측에 있어 중요한 요인을 찾기 위해 변수중요도를 확인해야 하며 변수중요도는 알고리즘 별 규칙에 의해 예측에 대한 변수별 영향도를 측정한 값이다.

본 논문에서는 Tree 기반의 알고리즘모형은 Feature Importance를 기본적으로 제공하기에 기본 모형을 사용했다. 따라서, 변수의 중요도를 정량적으로 측정할 수 있으며, 수치가 클수록 모델 결과를 예측하는데 더 큰 영향을 끼친다.

XGBoost를 인천광역시 침두 및 비침두 사고 심각도에 영향을 미치는 수치는 위의 Fig. 2와 같다. Fig. 3는 침두시간의 변수 중요도 상위 3개는 피해운전자 차종_승합, 가해운전자 차종_승합, 범규위반_차로위반으로 분석이 되었다. 비침두시간의 변수 중요도 상위 3개는 피해운전 차종_건설기계, 도로형태_단일로-터널, 도로형태_단일로-고가도로위-로 분석 되었다.

인적요인은 침두와 비침두 모두 가해자 성별, 가해자 연령, 피해자 성별로 구분되어 있으며, 침두시간에 가장 심각도에 큰 영향을 주는 요인은 가해자 성별이 남성으로 0.014939으로 분석되었다. 가해자 연령 중에는 60세 이상의 경우가 변수중요도 0.017846으로 높게 나타났다. 피해자 성별은 남성이 변수중요도가 0.014738 분석되었다. 비침두시간에 가장 심각도에 큰 영향을 주는 요인은 가해자 성별이 남성으로 0.006223으로 분석되었고, 가해자 연령 중 20대~30대가 0.010086으로 높게 분

석되었다.

시간 환경요인은 요일, 기상상태로 구분되어 있으며 가장 심각도에 큰 영향을 주는 요인은 침두시간은 주말로 0.015323으로 분석되었다. 기상상태로는 비가 오는 상황이며 0.011215로 분석되었다. 비침두시간의 심각도에 큰 영향을 주는 것은 주말로 0.00856으로 분석되었고, 기상상태는 맑음으로 0.012229로 분석되었다.

사고요인은 가해 운전자 차종, 피해운전자 차종, 사고유형, 법규위반, 도로형태로 구분되어 있다. 가해 운전자 차종에서는 승용차, 승합차, 화물차, 특수차, 건설기계, 원동기, 농기계, 이륜차, 자전거, 개인형이동수단, 사륜오토바이, 기타로 구분되어 있으며, 이 중 침두시간의 사고심각도에 큰 영향을 주는 요인은 승용차 0.015468로 분석되었고 그 다음은 건설기계인 0.012999로 분석이 되었다. 비침두시간의 사고 심각도의 큰 요인은 승합차로 0.015793으로 분석되었고 다음 심각도로는 화물차로 0.014282로 분석되었다. 피해운전자 차종에서 침두시간의 사고심각도는 승용차로 0.025563으로 분석되었고, 비침두시간의 사고심각도는 건설기계로 0.293379로 분석되었다,

사고유형은 침두시간은 차대차_정면충돌로 0.22879로 분석되었으며 비침두시간의 사고 심각도는 차대차_추돌로 0.017553으로 가장 크게 분석이 되었다.

법규위반의 사고 심각도는 과속, 교차로운행방법위반, 보행자 보호 의무 위반, 불법 유턴, 신호위반, 안전거리미확보, 안전운전 불이행, 중앙선침범, 직진우회전진행방해로 구분이 되어 있으며, 이 중 가장 침두시간 사고 심각도에 영향을 미치는 것은 차로위반으로 0.037152로 분석되었고, 비침두시간의 사고 심각도에 영향을 미치는 영향은 안전운전불이행으로 0.017955로 분석이 되었다.

도로형태는 터널, 고가도로 위, 교량, 지하차도 내, 횡단보도 내, 교차로안, 교차로 부근, 주차장, 기타로 분석이 되었으며 이 중 침두시간 사고 심각도는 단일로_교량에서 0.029339로 분석되었다. 비침두시간의 사고 심각도는 단일로_터널로 0.101021로 분석되었으며 교차로_횡단보도 내 0.010752로 분석이 되었다.

분석한 결과 가장 변수중요도가 높게 나온 변수는 사고유형 중 횡단 중일 때 가장 높은 값을 보였으며 피해자 성별이 남자, 사고유형 기타일 때, 야간시간대일수록 높게 나타난 것을 확인할 수 있었다.

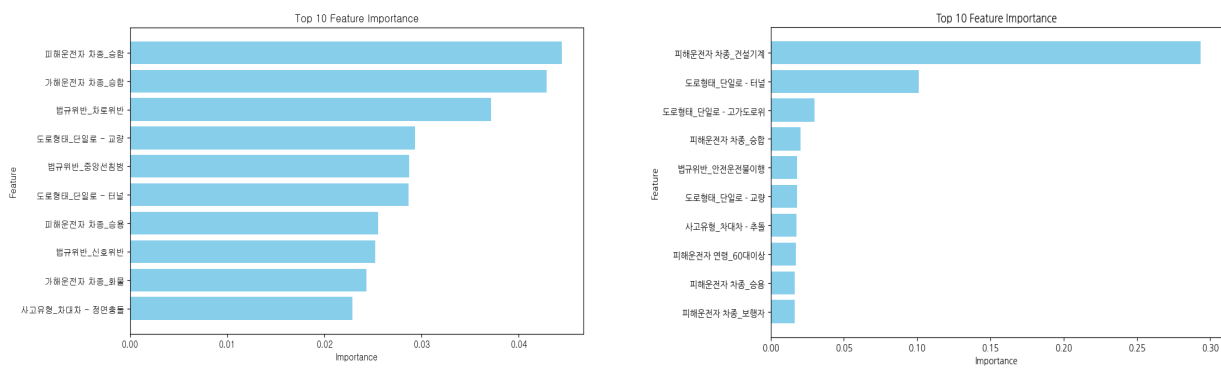


Fig. 2. Variable importance (peak), Variable importance (Non-peak)

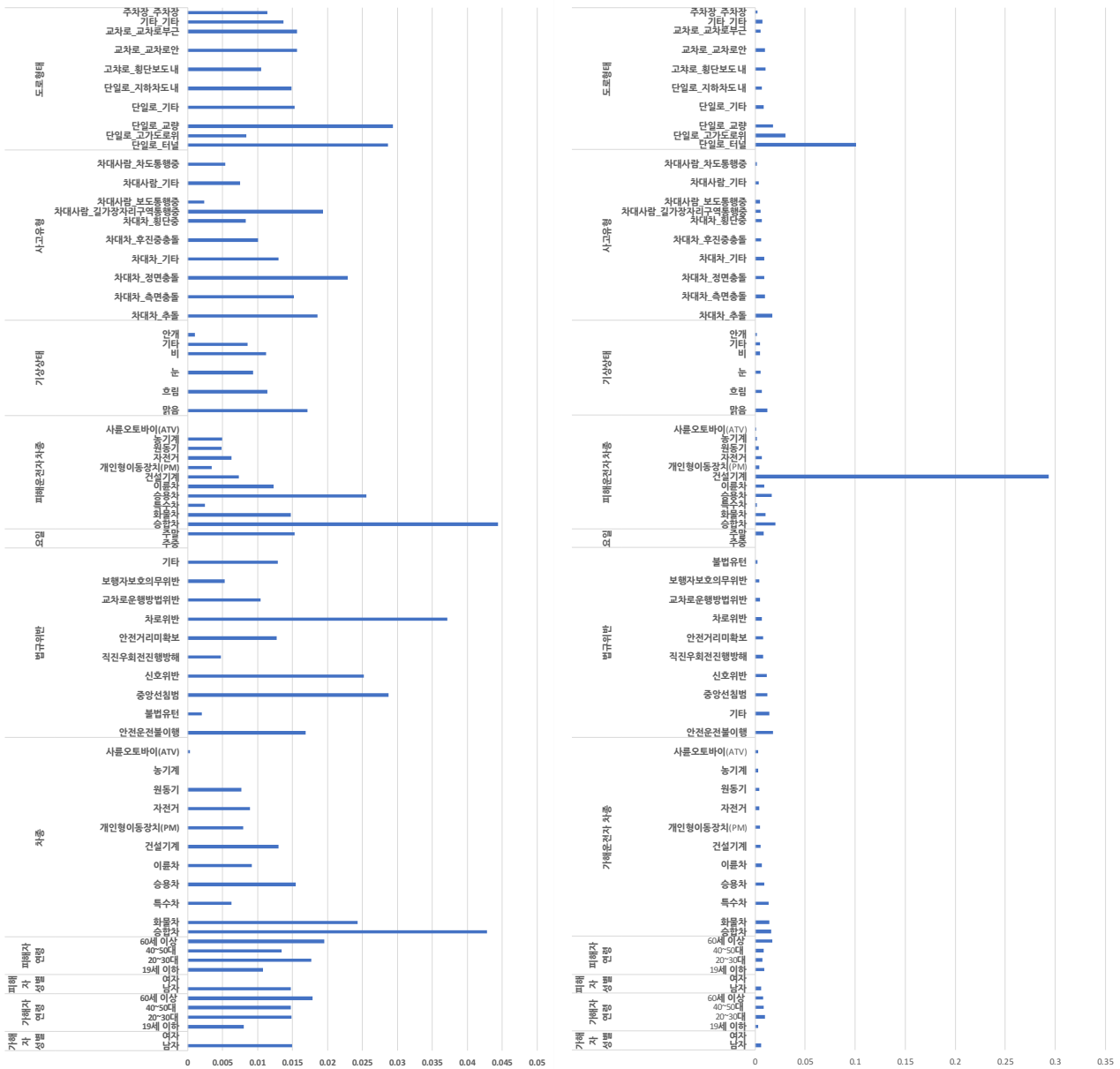


Fig. 3. Variable importance (peak), Variable importance (Non-peak)

결론

본 연구는 국내 자동차 보급 대수가 점차 증가 되고 있고, 삶의 질 향상에 따라 통행의 증가 등으로 자동차 통행량 또한 증가하고 있다. 자동차 통행량이 증가함에 따라 침두 시간과 비침두 시간의 교통사고에 분석을 위해 머신러닝을 활용하여 침두 시간의 교통사고와 비침두시간의 교통사고 모형을 구축하고자 했다. 인천광역시의 발생한 침두와 비침두 시간의 교통사고 심각도에 영향을 미치는 요인을 분석하고자 하였고, 침두시간과 비침두 시간의 교통사고 심각도의 차이를 분석하였다.

연구의 공간적 배경은 인천광역시로 설정했고 시간적 배경으로는 TAAS에서 제공하는 자료 2021년부터 2023년까지로

설정하였다. TAAS에서 제공하는 교통사고데이터를 가해자 인적요인, 피해자 인적요인, 도로형태, 사고유형, 요일, 노면상태, 기상상태, 법규위반을 활용하였다.

TAAS에서 구득한 자료를 XGBoost를 활용하여 인천광역시에서 발생한 침두시간 및 비침두 시간 교통사고 심각도 모형을 구축하였으며 XGBoost자체 모형을 활용하여 사고 심각도에 영향을 미치는 요인들을 분석하고 상위 10개의 변수중요도를 추출하였다.

인천광역시에서 발생한 침두 및 비침두 사고 심각도 모형의 상위 10개의 변수중요도를 도출한 결과, 피크시간인 침두시간에 사고 심각도가 높아지는 영향을 미치는 변수는 피해운전자 차종_승합, 가해운전자 차종_승합, 법규위반_차로위반, 도로형태_단일로 교량, 법규위반_중앙선침범, 도로형태_단일로-터널-, 피해운전자 차종_승용, 법규위반_신호위반, 가해운전자 차종_화물, 사고유형_차대차-정면충돌-으로 분석되었다. 침두시간의 사고심각도의 원인은 다음과 같이 판단했다.

침두시간의 교통사고 심각도는 출·퇴근 시간의 많은 교통량으로 인한 운전 피로도는 높아지고 집중도가 떨어지므로 법규위반을 하며 사고 심각도 높아지는 것으로 판단된다. 또한, 승합차와 화물차의 사고 및 피해가 많은 이유도 출·퇴근 시간에 집중되는 대형 차량 교통량으로 인하여 사고 심각도가 높아지는 것으로 판단된다.

비침두 시간의 교통사고 심각도가 높아지는 영향을 미치는 변수는 피해운전자 차종_건설기계, 도로형태_단일로-터널-, 도로형태_단일로-고가도로 위, 피해운전자 차종_승합, 법규위반_안전운전 불이행, 도로형태_단일로-교량-, 사고유형_차대차-추돌-, 피해운전자 연령_60대이상, 피해운전자 차종_승용, 피해운전자 차종_보행자로 분석되었다. 비침두시간의 사고심각도의 원인은 다음과 같이 판단된다.

건설기계 차량은 피크시간의 통행이 대부분 금지되기 때문에 피크 시간인 침두시간을 제외한 시간의 통행을 해야 한다. 따라서 건설기계 차량이 피크시간 때와는 다르게 변수가 높게 나온 것으로 판단된다. 피해운전자 연령이 60대 이상인 경우에는 은퇴를 한 고령층이 이동하는 경우가 많기에 피크시간과 다르게 사고 심각도가 높게 나온 것으로 판단된다. 또한, 도로형태의 사고심각도가 높은 이유는 피크 시간은 속도를 낼 수 없지만 논피크 시간에는 대부분 속도를 어느 정도 낼 수 있기에 도로형태에 따른 교통사고 심각도가 높게 나온다고 판단된다.

따라서 이러한 분석결과가 나온 원인 및 교통사고 심각도가 높아지는 것을 방지하기 위한 대책의 관한 연구와 침두시간의 많은 교통량을 분산할 수 있는 연구도 또한 필요할 것으로 판단된다.

References

- [1] Chen, T, Calos, G. (2016). "XGBoost: A scalable tree boosting system." In Proceedings of the 22nd Acm Sigkdd International Conference on Knowledge Discovery and Data Mining, New York, US, pp. 785-794.
- [2] Kwon, C.-W, Chang, H.-H. (2021). "Comparative analysis of traffic accident severity of tow-wheeled vehicles using XGBoost." Journal of Information Technology Services, Vol. 20, No. 4, pp. 1-12.
- [3] Lee, J.-J., Lee, Y.-R., Lim, D.-H., Ahn, H.-C. (2021). "A Study on the Employee Turnover Prediction using XGBoost and SHAP." Journal of Information Systems, Vol. 30, No. 4, pp. 21-42.
- [4] Lee, J.-J., Lee, Y.-R., Lim, D.-H., Ahn, H.-C. (2021). "A study on the employee turnover prediction using XGBoost and SHAP." Korean Society for Information System, Vol. 30, No. 4, pp. 21-42.
- [5] Lee, Y.-J., Sun, J.-W. (2020). "Predicting highway concrete pavement damage using XGBoost." Journal of Construction Engineering and Management, Vol. 21, No. 6, pp. 46-55.

- [6] Lee, Y.-J., Sun, J.-W. (2020). "Predicting highway concrete pavement damage using XGBoost." *Korean Journal of Construction Engineering and Management*, Vol. 21, No. 6, pp. 46-55.
- [7] Sim, D.-C., Lee, J.-Y., Jang, J.-W., Lee, M.-W. (2022). "Prediction of chloride concentration in groundwater on Jeju Island using XGBoost regression machine learning." *Journal of the Geological Society of Korea*, Vol. 58, No. 2.
- [8] Yoon, B.-J., Ko, E.-H, Yang, S.-R. (2016). "Comparative analysis on peak and non-peak hours traffic accident using logistic regression analysis." *The Korean Society of Disaster Information Regular Academic Conferences and Special Seminars*, Il San, Korea, pp. 283-284.
- [9] Yoon, B.-J., Lee, S.-M., Lwin, W.-Y. (2024). "Studying the comparative analysis of highway traffic accident severity using the random forest method." *Journal of the Society of Disaster Information*, Vol. 20, No. 1, pp. 156-168.
- [10] Yoon, B.-J., Lee, S.-Y., Jung, S.-Y. (2017). "A study on the factors of highway traffic accidents affecting the EPDO." *The Korean Society of Disaster Information*, Vol. 2017, No. 11, pp. 251-252.
- [11] Yoon, J.-H., Lee, S.-G. (2019). "Comparative analysis of factors affecting the severity of pedestrian crash by daytime and nighttime in Seoul, Korea." *Journal of Korea Planning Association*, Vol. 54, No. 7, pp. 70-88.