

# 머신러닝을 활용한 사회·경제지표 기반 산재 사고사망률 상대비교 방법론

김 경 훈\* · 이 수 동\*

\*울산대학교 산업안전보건전문학과

## Socio - economic Indicators Based Relative Comparison Methodology of National Occupational Accident Fatality Rates Using Machine Learning

Kyunghun Kim\* · Sudong Lee\*

\*Department of Occupational Safety and Health, University of Ulsan

### Abstract

A reliable prediction model of national occupational accident fatality rate can be used to evaluate level of safety and health protection for workers in a country. Moreover, the socio-economic aspects of occupational accidents can be identified through interpretation of a well-organized prediction model. In this paper, we propose a machine learning based relative comparison methods to predict and interpret a national occupational accident fatality rate based on socio-economic indicators. First, we collected 29 years of the relevant data from 11 developed countries. Second, we applied 4 types of machine learning regression models and evaluate their performance. Third, we interpret the contribution of each input variable using Shapley Additive Explanations(SHAP). As a result, Gradient Boosting Regressor showed the best predictive performance. We found that different patterns exist across countries in accordance with different socio-economic variables and occupational accident fatality rate.

**Keywords :** Occupational Safety and Health, Occupational Accident Fatality Rate, Machine Learning, Gradient Boosting Regressor, SHAP

### 1. 서론

대용량 데이터의 수집·저장·분석·활용에 관련된 하드웨어와 소프트웨어 기술의 발전은 생산 및 품질관리, 마케팅, 재난 예방, 의료·건강관리, 우주개발, 방위산업 등 여러 산업 전반에 큰 변화와 혁신을 가져왔고, 그 활용의 범위가 계속해서 넓어지고 있다. 산재 예방 분야에서도 수집된 데이터에 머신러닝(Machine Learning) 기법을 접목하여 예방역량을 향상하고자 하는 다양한 시도들이 이뤄지고 있다. 감독 당국 차원에서 한정된 인력이나 예산을 산재 예방 활동에 더 효과적으로 활용하기 위해 상대적으로 사고위험이 큰 업종·사업장을 특정하여 선제 대응하

려는 시도가 있었다. 자연어 인식·생성 기술을 접목하여 안전보건 예방정보나 법령정보를 쉽게 검색·제공하는 데 도움이 되는 산재 예방 데이터베이스 구축 시도도 있었다. 그 외, 사업장 내 실시간 영상 데이터를 분석하여 근로자의 위험 행동을 금지하거나[1], 근골격계질환 예방을 위한 올바른 작업 방법·자세를 제시하는 등의 다양한 산재 예방 노력이 시도되고 있다.

본 논문에서는 한 국가의 산재사고 수준을 예측하고 해석하기 위해 사회·경제적 발전 지표 데이터를 기초로 산재사고사망률을 예측하는 머신러닝 모델을 활용한 방법론을 제안하고자 한다. 단순 예측에 그치지 않고, 머신러닝 모델 해석 기법인 Shapley Additive Explanations

†Corresponding Author : Sudong Lee, Department of Industrial Engineering, University of Ulsan, 93 Daehak-ro, Nam-gu, Ulsan, Tel : 052-259-2174, E-mail: sudonglee@ulsan.ac.kr

Received November 17, 2022; Revision December 16, 2022; Accepted December 23, 2022

(SHAP)을 활용하여 국가의 산재사고사망률을 결정하는 사회·경제적 요인을 데이터 기반으로 해석할 수 있는 도구를 제안한다.

다양한 데이터를 통해 국가의 사회·경제적 발전 수준을 파악할 수 있다. 예를 들어, 국민 소득 수준, 경제 성장 속도 등 경제 성장 지표, 영아사망률이나 기대수명과 같은 보건·의료 발전 수준 지표, 산업별 취업자 수 및 실업률과 같은 고용지표, 경제의 역동성을 보여주는 국제통상 지표 등을 활용할 수 있다. 이러한 데이터들로부터 이들 지표와 국가의 산재사고사망률 간의 관계를 나타내는 예측 모델을 구현하여, 특정 국가의 사회·경제적 발전 수준 대비 산재 예방 수준의 적절성을 비교·평가할 수 있는 수단으로 활용할 수 있다.

## 2. 문헌 연구

최근 인공지능 기술을 재난 예방·의료·안전 분야에 접목하는 다양한 연구들이 선행되었다. 교통사고 예방을 위해 머신러닝 등 인공지능 기법을 적용한 연구로, Parsa 등[2]은 시카고 교통 당국에서 실시간 수집되는 통행량 데이터에 XGBoost와 SHAP을 적용해 사고 발생 탐지 및 영향 인자 평가를 시도하였다. Sheng 등[3]은 2014~2019 파키스탄 N5 고속도로의 교통사고 데이터를 4종류의 Gradient Boosting 기법으로 분석하여 LightGBM이 가장 높은 예측력을 보임을 확인하고, SHAP을 적용하여 각 입력 변수의 예측 기여도를 분석하였다.

의료분야에서는 환자 진료와 처방에 관련된 대용량 데이터 인프라가 상대적으로 빠르게 구축되었다. Wang 등[4]은 관상동맥류 심장질환 발병 환자의 3년 내 사망률을 머신러닝과 SHAP 기법을 활용해 분석하였다. Seki 등[5]은 2009~2017년 동경대학교 병원에 입원한 환자 119,160명의 25개 특성 데이터와 머신러닝 기법을 활용하여 입원 시 환자의 사망 가능성에 대한 위험성을 예측하고 사망을 방지하는 데 필요한 의료자원 수준을 예측하였다. Gao 등[6]은 COVID-19에 감염된 환자의 임상 데이터를 활용하여 감염 환자의 사망 위험성을 예측하는 머신러닝 모델을 학습하였다. Tseng 등[7]은 671건의 임상 사례로부터 환자 특성 데이터를 분석하여 심장 수술과 연관된 급성 신장 손상으로 인한 환자의 사망을 예측하는 머신러닝 모델을 학습하고 SHAP을 통해 각 변수의 기여도를 분석하였다. Li 등[8]은 372,808건의 전립선암 환자 데이터에 머신러닝 및 SHAP을 적용하여 발병요인의 비선형적 상호작용을 분석하였다.

상대적으로 도입은 늦었지만, 산재 예방 분야에서도 머

신러닝·딥러닝 등 인공지능 기법을 적용한 데이터분석과 모델링을 접목하려는 연구들이 활발하게 진행되고 있다. 머신러닝을 접목한 자연어 분석기법으로 관련 문건, 기술 기준이나 법령 등을 분석하여 유의미한 결과를 도출하려는 시도들이 있었다. Hegde 등[9]은 머신러닝 기반 문건 분석을 통해 산업별 위험성 평가를 수행하였고, Sattari 등[10]은 석유·가스업종 내 8,199건의 PSM (Process Safety Management) 보고서를 머신러닝과 Bayesian 기법으로 분석하여 사고 예방 전략의 유효성을 평가하였다. Baker 등[11]은 사고 관련 데이터와 문건을 머신러닝 기법으로 분석하여 건설 현장에서 발생하는 사고 정도, 종류, 상해 부위의 특성을 파악하였다. Pishgar 등[12]은 여러 산업의 산업안전 분야에 인공지능 기술을 접목하는 방안에 대한 문헌 고찰을 수행하였다.

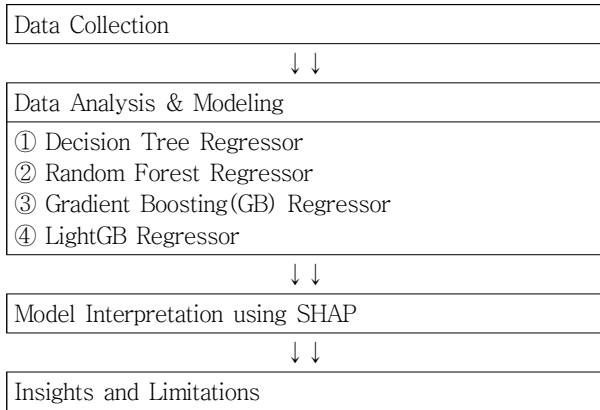
산재 사고의 예측 및 예방에 머신러닝 모델을 적용한 연구들도 있다. Koc 등[13]은 2012~2020년 터키에서 발생한 건설 현장의 산재사고 393,160건의 데이터로부터 사고 건수를 예측하기 위해 여러 머신러닝 기법을 비교하여 Hybrid Wavelet Artificial neural Network가 가장 우수한 예측력을 보이는 것을 확인하였다. Mohammed 등[14]은 미국 Occupational Safety and Health Agency (OSHA) 산재 데이터에 머신러닝 기법을 적용하여 산재 발생 여부를 예측하였는데, Gradient Boosted Trees의 예측력이 가장 높음을 확인하였다. Gondia 등[15]은 건설 산재 데이터 분석을 기반으로 사고 발생의 위험이 큰 장소와 공정을 예측하여 선제적으로 예방하는 연구를 수행하였다. 국내 연구 사례로 Oh 등[16]은 국내 사업장 약 5천 개의 특성 데이터에 4가지 머신러닝 기법을 적용, LightGBM이 가장 예측력이 우수함을 확인하고 SHAP을 적용하여 각 변수의 설명력을 확인하였다. Kim 등[17]은 국내 컨테이너 항구들의 물류 데이터와 사고 발생 데이터를 결합하여, 머신러닝 모델을 통해 사고를 예측하였다.

## 3. 연구 방법

### 3.1 분석 방법

머신러닝 기반 산재사고사망률 예측 모델을 학습하기 위해 국가의 사회·경제적 발전 수준 지표 데이터를 수집하였다. 두 번째 단계로, 수집된 지표들을 독립변수로, 산재사고사망률을 종속변수로 설정하고, 4가지 머신러닝 예측 모델을 학습한 후 비교 평가하였다. 셋째, 문헌 연구에서 살펴본 선행 연구 사례와 같이, 머신러닝 모델 해석 기법인 SHAP을 활용하여 학습된 예측 모델을 바탕으로 중

속변수의 예측에 대한 각 독립변수의 기여도를 분석하였다. 마지막으로 본 연구를 통해 발견한 시사점과 예측 모델 개선을 위해 고려할 사항을 제시하고자 한다. [Figure 1]은 본 연구 방법론을 단계별로 요약한다.



[Figure 1] Procedure of this study.

### 3.2 데이터 수집 및 변수 정의

예측 모델 학습을 위해 대한민국을 포함한 11개 선진국의 1991~2019년의 산재사고사망률과 17개 사회·경제적 지표 데이터를 수집하였다. 종속변수인 산재사고사망률은 <Table 1>과 같이 국가별 데이터 제공 기관에서 수집하였다. 독립변수인 사회·경제지표 데이터는 World Bank (<https://data.worldbank.org>)에서 제공하는 데이터를 활용하였다. 최초 수집된 5,878건의 데이터 중 대상 변수에 결측값이 있는 데이터를 제외한 5,706건의 데이터를 분석하였다.

<Table 1> Data source of occupational accident fatality rate. (11 countries, 1991~2019)

Country	Data Source
Europe*	WHO European Health Information
Korea	(Before 1998) KOSHA (After 1998) e-나라지표
U.S.A	① U.S. BUREAU OF LABOR STATISTICS (BLS) ② Center for Disease Control and Prevention (CDC)
Singapore	ILOSTAT

\* Europe : Denmark, Finland, France, Germany, Norway, Spain, Sweden, United Kingdom (8 countries).

분석에 사용된 국가별 사회·경제적 특성 지표의 정의 및 약어는 <Table 2>와 같다.

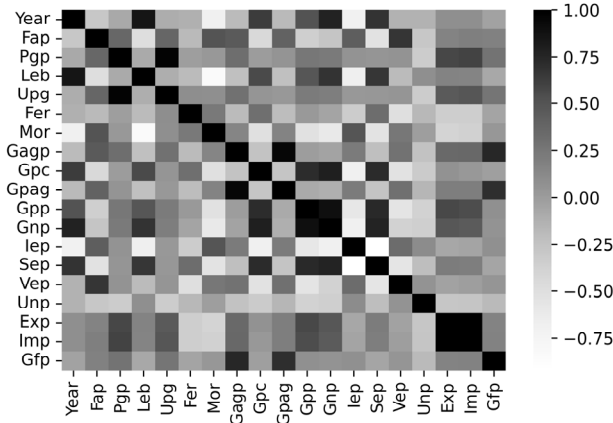
<Table 2> Socio-Economic indicators for machine learning models

Socio-Economic Indicators	Abbre-viation
Dependent Variable	
Accident fatality per 100,000 workers	Fap
Growth Indicator	
Population growth (annual %)	Pgp
Urban population growth (annual %)	Upg
GDP growth (annual %)	Gagp
GDP per Capita growth, Nominal (annual %)	Gpag
Growth fixed capital formation (annual % growth)	Gfp
Social Development Indicator	
Life expectancy at birth, total(years)	Leb
Mortality rate, under-5 (per 1,000 live births)	Mor
Fertility rate, total (birth per woman)	Fer
Income Level Indicator	
GDP per capita, Nominal	Gpc
GDP per Capita, PPP	Gpp
GNI per capita, PPP (current international \$)	Gnp
Employment Indicator	
Industry Employment % (ILO ISIC 2·3·4)	Iep
Service Employment %	Sep
Vulnerable Employment %	Vep
Unemployment %	Unp
International Commerce Indicator	
Export good and service (% of GDP)	Exp
Imports of goods and services (% of GDP)	Imp

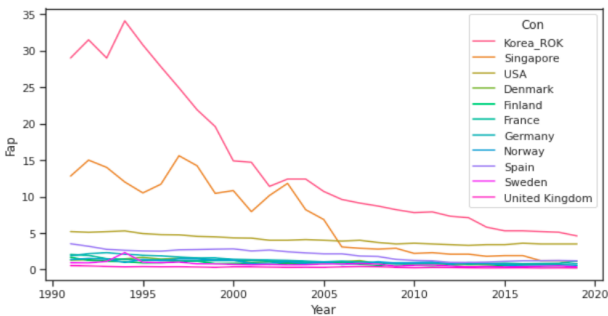
## 4. 데이터 분석 및 모델링

### 4.1 상관분석 및 종속변수 고찰

[Figure 2]는 각 변수의 Pearson 상관 분석 결과를 나타낸다. 비교대상 변수사이의 상관계수가 높을수록 그림에서 짙은 색으로 표시되었다. 독립변수로 선정된 사회·경제적 지표 중 산재사고사망률(Fap)과 높은 상관관계를 가지는 변수를 다수 확인할 수 있다.



[Figure 2] Heat-map of Pearson correlation analysis of applied variables.



[Figure 3] Fatality per 100,000 workers (11 countries, 1991~2019).

[Figure 3]은 분석대상 11개국의 연도별 산재사망사고 십만인률을 나타낸다. 영국 등 유럽 9개국과 미국의 산재사고사망률은 완만하게 감소하는 데 반해, 싱가포르와 대한민국의 경우 상대적으로 높은 수치로부터 급격한 하락세를 보이는 것을 확인할 수 있다. 싱가포르의 산재사고 사망률은 지속적으로 감소하여, 2006년도에는 미국보다 낮아지며 유럽의 수준과 비슷해지는 데 반해, 대한민국의 경우 하락세는 명확하나 2019년에도 타 국가에 비해 상대적으로 높은 수준에 머무는 것을 알 수 있다.

### 4.2 예측모델 학습

수집된 17개 사회·경제적 지표 독립변수로 종속변수인 산재사고사망률을 예측하기 위한 모델을 학습하였다. <Table 1>의 4가지 머신러닝 회귀분석 방법론을 적용하고 비교 분석하였다.

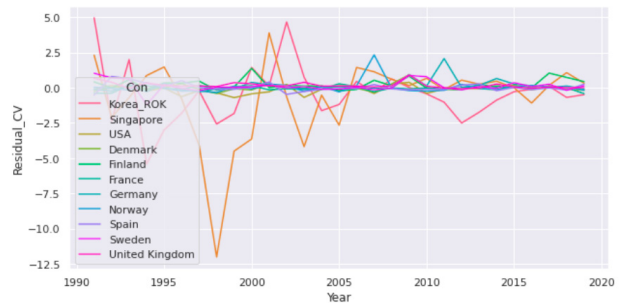
11개 국가가 5,706건의 데이터에서 국가별로 동수의 데이터가 추출되도록 균등 샘플링 하였다. 학습된 모델의 예측성능 검증을 위해 5겹 교차검증(5-fold cross validation) 수행 후, 학습(train) 및 테스트(test) 데이터에 대한 평균 절대오차(Mean Absolute Error, MAE)의 평균을 계산

하였으며, 그 결과는 <Table 3>과 같다. 4개의 머신러닝 방법론 중 GBR이 테스트 데이터 기준 평균 MAE 0.4818로 가장 우수한 예측성능을 보였다.

<Table 3> Evaluation Results of Models

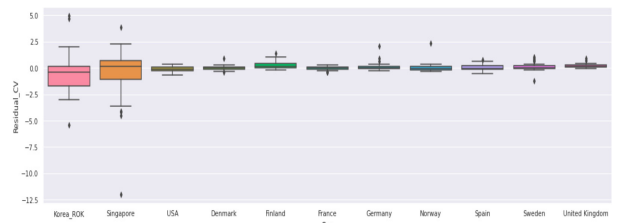
Regressor Models	MAE_train	MAE_test
GBR	0.0003	0.4818
LightGBM	0.0144	0.9089
Random Forest	0.2059	0.5971
Decision Tree	0.6505	1.0617

[Figure 4]는 위에서 학습된 GBR 모델의 연도별 잔차(residual)를 나타낸다. 다른 9개 국가에 비해 싱가포르와 대한민국의 잔차 크기가 큰 것을 확인할 수 있었다. 이를 통해, 해당 두 국가의 경우 다른 전체 국가의 데이터로 학습된 예측 모델이 공통으로 설명하지 못하는 산재사고 사망률 결정 요인이 존재한다는 사실을 유추할 수 있다.



[Figure 4] Residuals of the prediction model by year.

[Figure 5]의 11개 국가의 산재사고사망률 잔차 분포를 시각화한 박스플롯에서도 싱가포르와 대한민국의 산재사고사망률 분포가 다른 국가에 비해 큰 산포를 가지며 이상치 또한 많은 것을 확인할 수 있다.



[Figure 5] Box plots for residuals of the prediction model by country

### 4.3 SHAP을 활용한 예측모델 해석

각 변수가 수립된 모델에 영향을 미치는 기여도를 확인하고자, 4개의 모델 중 가장 우수한 성능을 보인 GBR에

의해 학습된 모델에 SHAP을 적용하였다.

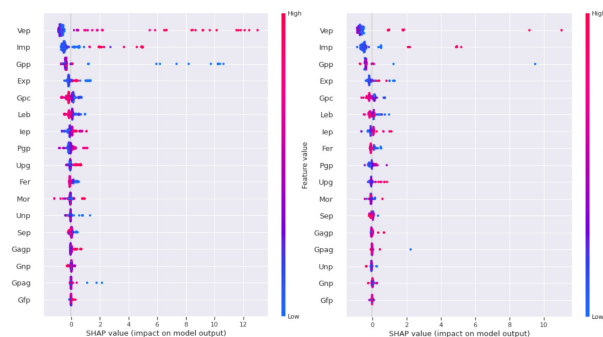
<Table 4> Importance of variables : SHAP analysis results using all data from 11 countries

Indicator	Importance	Indicator	Importance
Gpp	0.2958	Leb	0.0151
Vep	0.2886	Iep	0.0104
Gpc	0.1412	Mor	0.0073
Unp	0.0782	Sep	0.0060
Upg	0.0423	Gnp	0.0055
Imp	0.0393	Fer	0.0032
Exp	0.0386	Gfp	0.0020
Pgp	0.0245	Gagp	0.0018

<Table 4>는 각 변수의 SHAP value로 도출된 중요도를 나타낸다. 학습된 산재사고사망률 예측 모델에서 기여도가 높은 변수들은 다음과 같다.

- ① 구매력 기준 1인당 GDP(29.58%)
- ② 고용 취약계층 근로자 비율(28.86%)
- ③ 1인당 명목 GDP(14.12%)
- ④ 실업률(7.82%)
- ⑤ 연간 도시거주 인구 증가율(4.23%)
- ⑥ GDP 대비 상품 및 서비스의 수입 비율(3.93%)
- ⑦ GDP 대비 상품 및 서비스의 수출 비율(3.86%)

위 결과는 여러 변인 중 소득수준(①+③=43.70%), 고용지표(②+④=36.68%), GDP 대비 교역금액 비중(⑥+⑦=7.79%) 등이 GBR의 산재사고사망률 예측에 가장 큰 영향을 미침을 알 수 있다.



[Figure 6] SHAP analysis results for train (left) and test (right) data.

국가별로 예측 모델의 변수 중요도가 다르게 산출되는지 확인하기 위하여 개별 예측 모델 학습을 수행하였다. 분석 대상 11개 국가 중 산재사고사망률의 변화가 가장 크고 예측 정확도가 가장 낮았던 두 국가인 대한민국과 싱가포르, 2019년도 기준 산재사고사망률이 가장 낮았던

영국과 핀란드를 대상 국가로 선정하였다. 머신러닝 예측 모델은 전체 국가 데이터에서 가장 높은 예측성능을 보인 GBR을 사용하였다. <Table 5>는 학습된 GBR 모델로부터 SHAP 변수 기여도를 산출하여, 기여도 기준 상위 5개 독립변수와 각각의 기여도를 요약하였다.

<Table 5> Variable importance: idiographic SHAP analysis results for Korea, Singapore, UK, and Finland

Korea		Singapore	
Indicator	Importance	Indicator	Importance
Sep	0.3245	Sep	0.3297
Gnp	0.2121	Gnp	0.2341
Leb	0.1831	Gpc	0.2306
Gpc	0.1152	Imp	0.0492
Iep	0.0551	Vep	0.0385
United Kingdom		Finland	
Indicator	Importance	Indicator	Importance
Iep	0.2765	Sep	0.3997
Mor	0.1722	Gnp	0.2035
Gpag	0.1289	Leb	0.1068
Sep	0.0953	Gpc	0.1028
Imp	0.0891	Iep	0.0579

대한민국과 싱가포르의 SHAP 분석 결과를 살펴보면, 서비스 고용 비율, 구매력 기준 1인당 GNI, 1인당 명목 GDP가 산재사고사망률에 가장 큰 영향을 미친 것을 확인할 수 있다. 이 결과를 통해 급격한 경제 성장을 이룬 대한민국과 싱가포르에서 3차 산업(서비스·금융 등) 비중과 국민 소득이 증가하면서 산재사고사망률이 감소하였음을 유추할 수 있다. 또한, 두 국가의 기여도 상위 5개 지표 중 3개가 서로 일치하여, 상호 유사한 사회·경제적 변인이 두 나라의 산재사고사망률에 영향을 미쳤다고 할 수 있다.

반면, 영국과 핀란드의 경우는 상위 5개 지표 중 서로 중복되는 것이 2개(건설·제조업 및 서비스 고용률)뿐이며, 그 비중의 차이도 크다. 핀란드는 대한민국, 싱가포르의 상위 5개 지표와 중복되는 지표가 3개인데 반해, 영국은 대한민국과 Sep, Iep, 싱가포르와는 Sep, Imp 각각 2개 지표만이 같음을 알 수 있다. 또한, 흥미로운 점은 전체 11개국 데이터를 분석한 SHAP 결과인 <Table 4>의 상위 5개 지표 중 <Table 5>의 개별 4개국 SHAP 분석 상위 5개 지표와 중복되는 것이 Gpc, Vep 2개 지표뿐이라는 것이다. <Table 4>의 11개 전체 데이터로 분석한 산재사고사망률 기여도와 대한민국의 결과를 비교하면 상위 5개 지표 중 중복되는 지표는 1인당 명목 GDP 1개에 불과하였다.

## 5. 결 론

### 5.1 산재예측 모델 구현과 활용

본 연구에서는 선진 11개국의 29년간의 사회·경제발전 지표 데이터에 머신러닝 기법을 적용하여 산재사고사망률 예측 모델을 학습하였다. 전체 11개국 데이터로 학습한 예측 모델의 SHAP 변수 기여도를 분석한 결과, 소득수준·고용지표·교역 비중이 상대적으로 산재 발생에 영향을 큰 영향을 미치는 것을 확인하였다.

다만, 대한민국, 싱가포르, 영국, 핀란드의 4개 개별국가 데이터로 각각 학습한 예측 모델의 결과와 비교 시, 국가별로 중요 변수가 상이하게 도출되는 것을 확인할 수 있었다. 이는 산재사고 사망 예측 모델 수립 시, 사회·경제적 특성이 유사한 국가들을 제한적으로 묶어서 모델을 구현하거나 개별국가 데이터만으로 모델을 학습할 때 예측 성능을 높일 수 있음을 의미하며, 선행하여 실시한 잔차분석 결과와도 부합한다.

### 5.2 시사점과 토론 의제

[Figure 3]의 산재사고사망 십만인률 그래프에서, 유럽 등 서구국가들은 산업혁명 이후 2백여 년 동안 산재사고사망률이 꾸준히 낮아져 지난 29년이 장기 시계열 하락 추이의 최종단계(saturated status)를 보였다 할 수 있다. 반면, 산업화 역사가 상대적으로 짧고 산재 예방 인프라가 늦게 형성된 싱가포르나 대한민국의 산재사고사망률은 상대적으로 큰 변화폭을 보인 것을 확인할 수 있다. 이 고찰은 산재사고사망률의 정확한 예측과 유의미한 국가 간 비교를 위해서는 산재 예방의 역사, 안전보건 수준 발전단계, 산업구조, 노동 인력구성 등 다양한 사회·경제적 차이를 고려해야 한다는 것을 시사한다.

<Table 4>와 <Table 5>를 비교·분석한 결과는 국가별 산업구조·노동인력구조·소득수준 등 특성 차이로 인해 사회·경제지표 데이터가 산재사고사망률 변화에 영향을 미치는 정도가 다를 수 있다. 이 결과는 모든 국가에 적용할 수 있는 단일 산재사고사망률 예측 모델을 구현하는 것이 적절하지 않으며, 사회·경제적 유사성이 크고 비슷한 발전단계에 있는 국가들의 데이터만을 묶어서 모델을 수립하는 것이 바람직함을 제시한다.

만약, 충분한 장기 시계열 데이터를 확보할 수 있다면 단일 국가별 예측 모델 수립을 통해 예측성능을 개선할 수 있을 것이다. 이런 단일 국가 모델의 잔차 분석 결과는 특정년도의 산재예방사업의 성과를 평가하는 수단으로도 사용할 수 있을 것이다.

### 5.3 연구의 한계 및 향후 연구과제

머신러닝 기반 예측의 정확도에 가장 큰 영향을 미치는 요인은 데이터의 질(quality)과 양(quantity)이다. 본 연구에 사용한 11개국 데이터 외에 아시아·아프리카·남아메리카 대륙 개발도상국의 산재 데이터를 포함한 보다 많은 데이터를 분석할 수 있었다면 더 많은 시사점 도출이 가능할 수 있겠으나, 신뢰할 만한 데이터를 확보하는데 어려움이 컸다.

또한, 산업별 생산 비중, 고용인력 비율, 고령(55세 이상) 노동자 비율, 외국인 노동 인력 비율, 국가별 평균 근로시간, 기타 노동 강도를 나타내는 지표 등 보다 세분된 독립변수를 활용한다면 예측 모델의 설명력을 높일 수 있을 것이다. 산업 구분 측면에서도 산업별 GDP 기여 비율, 생산·노동 인력구조의 변화 등 폭넓은 사회·경제지표 데이터를 수집하여 국가별 산재 예측 모델을 구현하는 것은 흥미로운 연구주제가 될 것이다. 본 연구의 결과가 대한민국을 포함한 각국의 산재 예방정책의 성과 평가 기준으로서 그 활용성이 높을 것으로 기대한다.

## 6. References

- [1] S. J. Choi, K. H. Jung(2021), "Statistical analysis of major accident reports and development of a real-time detection model for portable ladder and safety helmet." *KSMS*, 23(1):9-15.
- [2] A. B. Parsa, A. Movahedi, et al.(2020), "Toward safer highways, application of XGBoost and SHAP for real-time accident detection and feature analysis." *Accident Analysis & Prevention*, 136.
- [3] D. Sheng, A. Khatkhat, et al.(2022), "Predicting and analyzing road traffic injury severity using boosting-Based ensemble learning models with shapley additive explanations." *International Journal of Environmental Research and Public Health*, 19(5):2925.
- [4] K. Wang, J. Tian, C. Zheng, et al.(2021), "Interpretable prediction of 3-year all-cause mortality in patients with heart failure caused by coronary heart disease based on machine learning and SHAP." *Computers in Biology and Medicine*, 137.
- [5] T. Seki, Y. Kawazoe, et al.(2021), "Machine learning-based prediction of in-hospital mortality using admission laboratory data: A retrospective,

- single-site study using electronic health record data." PLoS One, 16(2):e0246640.
- [6] Y. Gao, G. Y. Cai, et al.(2020), "Machine learning based early warning system enables accurate mortality risk prediction for COVID-19." Nat Commun, 11:5033.
- [7] P. Y. Tseng, Y. T. Chen, et al.(2020), "Prediction of the development of acute kidney injury following cardiac surgery by machine learning." Crit Care, 24:478.
- [8] R. Li, A. Shinde, et al.(2020), "Machine learning - based interpretation and visualization of nonlinear interactions in prostate cancer survival." JCO Clinical Cancer Informatics, 4:637-646.
- [9] J. Hegde, B. Rokseth(2020), "Applications of machine learning methods for engineering risk assessment - A review." Safety Science, 122.
- [10] F. Sattari, F. Macciotta, et al.(2021), "Application of Bayesian network and artificial intelligence to reduce accident/incident rates in oil & gas companies." Safety Science, 133.
- [11] H. Baker, M. R. Hallowell, et al.(2020), "AI-based prediction of independent construction safety outcomes from universal attributes." Automation in Construction, 118.
- [12] M. Pishgar, S. F. Issa, et al(2021), "A novel framework to review artificial intelligence and its applications in occupational safety and health." Int J Environ Res Public Health, 18(13):6705.
- [13] K. Koc, Ö. Ekmekcioğlu, et al.(2022), "Accident prediction in construction using hybrid wavelet-machine learning." Automation in Construction, 133.
- [14] J. Mohammed, M. J. Mahmud(2020), "Selection of a machine learning algorithm for OSHA fatalities." TEMSCON, 1-5.
- [15] A. Gondia, M. Ezzeldin, et al.(2022), "Machine learning - based decision support framework for construction injury severity prediction and risk mitigation." ASCE, 8(3).
- [16] H. R. Oh, A. L. Son, et al.(2021), "Occupational accident prediction modeling and analysis using SHAP." Journal of Digital Contents Society, 22: 1115-1123.
- [17] J. H. Kim, J. E. Kim, et al.(2021), "Machine learning-based models for accident prediction at a Korean container port." Sustainability, 13(16): 9137.
- [18] J. Takala, et al.(2014), "Global estimates of the burden of injury and illness at work in 2012." J. Occup Environ Hyg. 11(5):326-37.

## 저자 소개



### 김 경 훈

현 울산대학교 산업안전보건전문학과 박사과정 재학 중  
현 산업안전보건공단 재직 중  
연세대학교 경영전문대학원 석사  
관심분야 : 산업안전보건, 기계안전



### 이 수 동

현 울산대학교 산업경영공학부 조교수.  
포항공대 산업경영공학과 박사.  
관심분야 : 산업인공지능, 데이터분석