

건설현장 안전사고 위험도 예측 모델 개발

Development of a safety accident risk prediction model at construction sites

○석 병 민* 김 상 용**
Seok, Byeong Min Kim, Sangyong

Abstract

Existing safety management methods rely heavily on the subjective judgment of safety managers, which results in a lack of consistency and makes rapid response difficult. In particular, safety accidents that occur at construction sites are often difficult to predict, making it difficult to prevent safety accidents. Because of this, there are limits to improving safety levels and reducing the possibility of accidents. To solve these problems, this study seeks to develop a model that quantitatively predicts the risk of safety accidents occurring at construction sites. Through this study, expect to be able to establish more objective and rapid safety management measures.

키워드 : 건설사고, 안전관리, 위험도 예측, 앙상블 모델

Keywords : Construction Accident, Safety Management, Risk Prediction, Ensemble Model

1. 서론

1.1 연구의 배경 및 목적

최근 산업재해 현황 분석에 따르면, 2022년 산업재해 발생 건수가 전년 대비 6.22% 증가했으며, 특히 건설업은 타 산업보다 사망자 비율이 상당히 높은 것으로 나타났다(KOSHA, 2022). 건설현장에서 발생하는 재해는 다른 산업군에 비해 강도 높은 재해가 발생할 가능성이 크기 때문에 근본적으로 예방할 수 있는 방법이 필요하다(Cho, Kim & Shin, 2017). 안전사고 예방 및 대응의 중요성이 강조되고 있지만, 현재 다수의 건설현장에서 안전관리가 관리자의 주관적 판단에 의존하고 있다. 사건 분석은 기록의 품질에 크게 좌우된다. 기록상의 문제와 사건 데이터의 이질성으로 인해 속성 간의 관계를 분석하는 과정에서 충돌이 발생할 수 있다(Lee et al, 2020). 이러한 문제로 일관성이 부족하고 안전사고의 예방 가능성이 달라질 수 있다. 과거 경험이나 주관적인 판단에 의존하지 않고 과학적이고 정량적인 분석을 통해 위험을 예측할 수 있는 방안이 필요하다. 따라서 본 연구는 과학적 분석을 기반으로 건설현장 안전사고 위험요인별 상관관계를 분석하고, 위험도 예측 모델을 개발한다.

1.2 연구의 범위 및 절차

연구의 프로세스는 그림1과 같고, 건설공사 안전관리 종합정보망(CSI)에서 수집된 데이터를 기반으로 한다. 데이터셋의 분석을 통해 필수 요인 항목을 선별하고, 전처리 과정을 거쳐 데이터의 질을 향상시킨다. 위험도 예측 모델의 학습을 위해 데이터는 훈련 데이터 70%와 테스트 데이터 30%로 나누었고, XG Boost 알고리즘을 활용한 기계 학습을 수행한다. 최종적으로, 본 연구의 목적은 모델의 예측 성능을 검증하고 위험요인별 관계를 분석하여 위험도 예측 모델 개발에 있다.

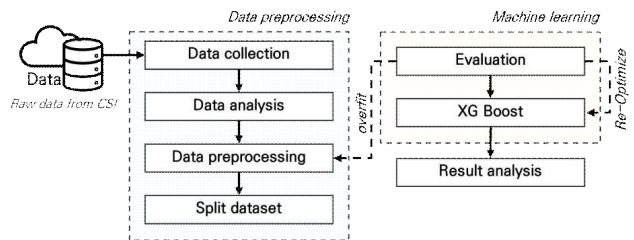


그림1. 연구의 프로세스

2. 문헌고찰

이전 연구들에서 다양한 머신러닝 기법이 산업 안전 예측 모델에 적용되었다. Kang & Ryu(2019)는 사고 조사 데이터와 기상데이터를 연계한 랜덤 포레스트 기법을 적용하여 71%의 예측 정확도를 보였고, 나이, 직업상 부상, 회사 규모, 공사비, 근무일 손실, 건축유형, 직업 등 21개의 변수를 사용하였다.

* 영남대학교 건축학부 건축공학전공 학사과정

** 영남대학교 건축학부 부교수, 공학박사

(Corresponding author : sangyong@yu.ac.kr)

이 성과는 정부(과학기술정보통신부)의 재원으로 한국연구재단의 지원을 받아 수행된 연구임(No.2020R1A2C1005263).

Luo, et al(2023)은 붕괴사고 산업재해 심각도 예측을 위해 랜덤 포레스트 모델을 사용하여 82%의 정확도를 보였다.

본 연구는 텍스트 기반의 범주형 데이터를 수치형 데이터로 전환하는 전처리 과정을 포함하여 앙상블 기법 기반의 XG Boost 모델을 개발함으로써 기존의 트리 기반 접근법에 비해 예측 정확도와 모델 신뢰성을 향상시키고자 한다.

3. 예측 모델 개발

3.1 데이터 설명

본 연구에서 사용한 데이터는 건설공사 안전관리 종합정보망(CSI)에서 2010년부터 2021년까지 발생한 총 9,290건의 안전사고 정보이다. 위험도 예측을 위한 필수 입력 값은 건설현장의 특성, 작업 환경요인, 그리고 작업자들의 특성이다. 또한, 안전관리 시스템의 구축 여부에 따라 안전사고 발생 위험도에 직접적인 영향을 미치므로 안전관리계획 여부도 고려해야 한다. 비정형 데이터인 사고명, 시설물, 사고 경위, 사고 발생 후 조치사항 등의 요인은 관리자의 주관적인 특성을 나타내기 때문에 이들 요인은 위험 요인에서 제외하고 분류하였다.

3.2 데이터 특성 분석 및 모델 결정

XGBoost(Extreme Gradient Boosting)는 체계적인 앙상블 기법을 통해 다수의 약한 예측 모형들을 결합하여 예측력을 극대화하는 고급 그래디언트 부스팅 알고리즘이다. 분류한 데이터셋은 범주형 데이터와 수치형 데이터가 혼합되어 있다. 모델의 학습 및 예측 성능을 높이기 위해 모두 수치형 데이터로 변환하는 전처리 작업을 거쳤다. 데이터 분석을 통해 위험 요인을 선정한 결과 ‘공중, 공사비, 공정률, 작업자수, 안전관리계획, 사고인지 시간, 기온(°C), 습도(%), 안전방호, 개인보호, 인적사고, 사고객체, 부위, 피해금액, 사고신고사유’ 총 16가지가 분류되었다.

4. 연구결과

4.1 모델의 성능 평가

평가 지표를 활용한 XGBoost 모델의 예측 성능 및 오차는 표1과 같다. 모델의 정확도(R1_Score)는 81%이며 MAE(Mean Absolute Error), MSE(Mean Squared Error), RMSE(Root Mean Squared Error)와 같은 오차 지표들은 낮은 값을 나타내어 모델의 예측 오차가 최소화 되었음을 나타낸다.

표1. 모델 오차 검토 및 예측 성능

Data	MAE	MSE	RMSE	R1_Score
Test data	0.0670	0.0258	0.1607	0.8191

4.2 결과해석

데이터의 위험요인별 상관관계에 대한 확률을 시각적으로 표현하기 위해 히트맵(Heatmap)을 활용하였고, 결과는 그림1과 같다. 그림1을 보면 Damage cost(피해금액)과 Risk intensity(위험강도)는 0.78의 높은 양의 상관관계를 보여, 피해 금액이 높아질수록 위험 강도가 증가할 가능성이 크다는 것을 나타낸다.

Construction type(공중)은 대부분의 다른 변수들과 낮은 상관관계를 보이고 있다. 이를 바탕으로 어떤 변수들이 다른 변수들에 영향을 더 많이 주는지, 또는 어떤 변수들이 독립적인지에 대한 상관관계를 확인하였다.

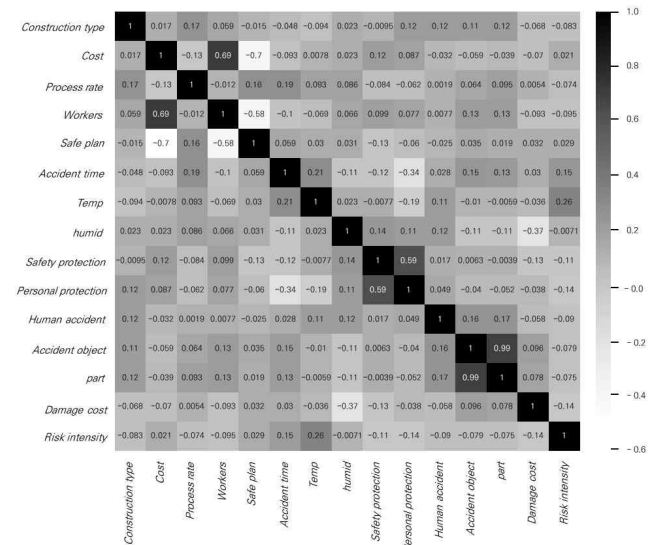


그림2. 위험요인별 상관관계 히트맵

5. 결론

본 연구는 CSI에서 수집한 안전사고 데이터셋을 분석 및 전처리 과정을 거쳐 건설현장의 안전사고 위험도에 대한 예측 모델을 개발하였다. 정제된 데이터를 학습하여 XG Boost 분석을 하였고, 모델의 예측 성능은 81%를 가진다. 히트맵 도구를 활용하여 데이터의 위험요인별 상관관계를 시각적으로 분석할 수 있었다. 이를 통해 모델에 다양한 상관관계를 고려하여 안전사고의 발생 가능성을 수치화하고 예측할 수 있으므로, 안전관리의 효율성을 높이고 중대사고의 발생 가능성을 줄일 수 있을 것으로 기대된다.

참고문헌

- Kang, K. & Ryu, H. (2019). Predicting types of occupational accidents at construction sites in Korea using random forest model. Safety Science, 120, 226-236.
- Ayhan, B. U. & Tokdemir, O. B. (2020). Accident analysis for construction safety using latent class clustering and artificial neural networks. Journal of Construction Engineering and Management, 146(3), 04019114.
- Luo, X., Li, X., Goh, Y. M., Song, X., & Liu, Q. (2023). Application of machine learning technology for occupational accident severity prediction in the case of construction collapse accidents. Safety science, 163, 106138.
- 조예림, 김연철, & 신윤석. (2017). 의사결정나무기법을 이용한 건설재해 사전 예측모델 개발. 한국건축시공학회지 (JKIBC), 17(3), 295-303.