

PA기법을 이용한 건축공사 공종별 사망사고 예측모델 개발에 관한 연구 - 의사결정나무를 중심으로 -

최정원¹ · 김한수^{2*}

¹세종대학교 건축학과 석사과정 · ²세종대학교 건축학과 교수

Predictive Analytics Model for Death Accidents in Building Projects by Trade - Based on Decision Tree -

Choi, Jeong Won¹, Kim, Han Soo^{2*}

¹Graduate Student, Department of Architecture, Sejong University

²Professor, Department of Architecture, Sejong University

Abstract : Compared with other industries, construction industry shows a higher rate of death accidents and recently companies' legal responsibilities are to be increasingly enforced. The trend causes tremendous concerns for construction firms and increases the importance of forecasting and pro-actively managing death accidents in construction fields. The objective of the study is to develop a predictive analytics model for forecasting death accidents in building projects based on a decision tree technique, which enables to forecast the probabilities of death accidents by trade. The use of the model helps to decrease risks of legal punishments and to assist the safe execution of building projects by forecasting and pro-actively managing death accidents.

Keywords : Predictive Analytics, Decision Tree Model, Construction Industrial Accidents, Construction Death Accidents

1. 서론

1.1 연구의 배경 및 목적

건설업은 타 산업에 비해 높은 산업재해율을 보이고 있으며, 사망사고를 포함한 산업재해의 발생은 건설사에게 피해자 보상, 공기지연, 법적 처벌 등에 따른 경제적 손실과 기회비용의 증가를 가져오고 있다(Bae & Lee, 2018). 최근에는 2022년 1월 시행을 앞두고 있는 '중대재해 처벌 등에 관한 법률(중대재해처벌법)'을 통해 법적 책임과 처벌이 강화되었으며 특히 사망사고에 대한 처벌 수준이 높은 것으로 나타나고 있다.

중대재해처벌법에 따르면 안전 및 보건확보 의무를 위반하여 1명 이상의 사망자가 발생한 경우, 사업주 또는 경영책임자 등에게 1년 이상의 징역 또는 10억원 이하의 벌금형을

부과하는 것으로 규정하고 있다. 중대재해처벌법은 건설사에게 큰 부담이 되고 있으며 이는 사망사고에 대한 사전 예측과 선제적인 대응의 필요성을 증가시키고 있다.

본 연구의 목적은 PA (Predictive Analytics)기법 중 하나인 의사결정나무(decision tree)를 활용하여 건축공사에서 발생할 수 있는 공종별 사망사고를 사전에 예측할 수 있는 모델을 개발하는데 있다. 본 연구에서 의미하는 공종별 사망사고 예측모델이란 건축공사의 조건에 따라 공종별로 발생할 수 있는 사망사고의 가능성(확률)을 예측하는 모델을 의미한다. 예측모델의 활용을 통한 사망사고에 대한 사전 예측과 선제적 대응은 법적 처벌을 예방하고 건축공사의 성공적인 수행이라는 측면에서 중요한 의의를 지닌다.

1.2 연구의 범위 및 방법

본 연구는 건설공사 중 건축공사에서 발생하는 다양한 산업재해 중 가장 치명적이며 중대산업재해로 분류되고 있는 사망사고를 연구범위로 설정하였다.

본 연구에서는 '건설공사 안전관리 종합정보망(www.csi.go.kr)'을 통해 2019년 7월~2021년 2월까지의 건축공사 사

* Corresponding author : Kim, Han Soo, Department of Architecture, Sejong University, Korea

E-mail: hskim@sejong.ac.kr

Received July 12, 2021 : revised July 27, 2021

accepted August 10, 2021

망사고 124건의 사례를 수집하여 데이터로 활용하였으며, 의사결정나무를 활용한 공종별 사망사고 예측모델의 구축을 위해 'IBM SPSS Statistics 27'을 사용하였다.

본 연구의 주요 단계는 다음과 같다.

(1) 문헌 조사를 통해 고용노동부의 건설 산업재해 현황과 PA기법 및 의사결정나무에 대해 조사하고, 건설 산업재해에 관한 선행 연구를 분석하였다.

(2) 건설 사망사고 사례 수집을 통해, 모델 구축을 위해 필요한 건축공사 사망사고 데이터를 수집하였으며, 분석 요인들을 선정하고 모델 구축 기준을 설정하였다.

(3) 'IBM SPSS Statistics 27'를 사용하여 의사결정나무 기반의 건축공사 공종별 사망사고 예측모델을 구축하였다.

(4) 구축된 예측모델을 활용하여 건축공사 공종별 사망사고를 예측하는 방법을 제시하고, 예측모델의 적합성을 평가하였다.

2. 건설 사망사고 및 의사결정나무

2.1 건설 사망사고 및 영향

고용노동부 자료에 따르면, 건설업은 타 산업에 비해 높은 사망사고 현황을 보이고 있다(MEL, 2021). 2020년 기준으로 건설업의 사망자수는 567명(1위), 제조업 469명(2위)로 나타났다. 또한, 근로자 10,000명당 발생하는 사망자 수의 비율을 나타내는 사망만인율 측면에서 보면, 2020년 건설업이 2.48‰, 제조업이 1.17‰로 나타났으며 이는 전(全) 산업 평균인 1.09‰에 비해서도 2배 이상의 높은 수준을 보이고 있다. 특히 주목해야 할 것은 사망만인율의 증가이며, 전년도 대비 2020년의 사망만인율 증감률은 건설업 19.2%P, 제조업 -4.1%P로 나타났다. 이와 같이 건설업의 사망사고 현황은 타 산업에 비해 심각하며, 건설사에게 다양한 측면에서 부정적인 영향을 끼치게 된다.

비용 측면에서 산업재해(사망사고)는 건설사에게 직접적인 영향을 주고 있다. 중대재해처벌법 제15조에 따르면 법에서 정한 의무를 위반하여 중대재해를 발생하게 한 경우 중대재해로 손해를 입은 사람에 대하여 그 손해액의 5배를 넘지 아니하는 범위에서 배상책임을 지며, 산업안전보건법 제55조에 의해서 중대재해 발생 시 고용노동부장관의 작업 중지 조치가 발생하여 공사 중지가 될 수 있다. 이로 인해 직접비용(예: 배상책임 비용)과 간접비용(예: 공사 중지에 따른 추가 비용)이 발생할 수 있다.

재해 유형 중 추락재해의 경우, 건설사의 산업재해 손실 비용은 추락재해 한 명당 직접비용은 약 1천9백만원이고 간접비용은 약 5천7백만원으로, 건설사 측면에서 추락재해 손실 비용은 부상자 한 명당 약 7천6백만원으로 추산되고 있

다(Sa et al., 2012). 사망사고 발생 시에는 추가적으로 위로금, 보상손실 등도 발생하는 것으로 나타난다(Yang & Kim, 2014). 즉, 산업재해(사망사고)는 직접비용 및 간접비용의 발생 측면에서 건설사에게 직접적인 영향을 주는 것이다.

산업재해(사망사고)는 건설사에게 비용이라는 직접적인 영향 외에도 각종 심사·자격의 기준 측면에서 간접적으로 부정적 영향도 끼치게 된다. 공사계약 종합심사낙찰제 심사기준(기획재정부계약예규 제544호)에 따르면, 사고사망만인율은 종합심사낙찰제의 사회적 책임 세부 심사기준에 해당된다. 또한, 산업안전보건법 시행규칙에 따르면 시공능력평가 시, 평균재해율의 1배 이상 2배 이내의 재해를 발생시킨 건설사업자에 대해서는 공사실적액의 감액이 발생하며, 국가계약법 시행령에 따르면 안전·보건 조치를 소홀히 하여 근로자 등에게 사망 등 중대한 위해를 가한 경우 입찰참가 자격이 제한된다. 즉, 산업재해(사망사고)는 법적 기준으로 인해, 심사·자격에 영향을 주는 것으로 나타나며, 이 역시 건설사에게 미치는 파급력이 크다고 할 수 있다.

상기에 제시된 바와 같이 사망사고의 발생은 건설사에게 심각한 현안이 되며, 사망사고의 발생을 줄이기 위한 다각적인 노력이 필요하고, 사망사고를 사전에 예측하고 선제적으로 대응하는 것은 이러한 노력의 일환이 될 수 있다.

2.2 PA기법 및 의사결정나무 고찰

2.2.1 PA기법

PA기법이란 과거·현재를 분석하여 미래를 예측하는 예측분석을 위해 활용되는 기법 또는 도구를 통칭하는 용어이다(Mishra & Silakari, 2012). 예측분석이란, 미래에 발생할 일에 대한 최적의 의사결정을 도출하기 위해, 과거 또는 현재의 데이터를 통해 미래를 예측하는 것이다(Baek, 2017). 예측분석은 상당 기간 중요한 주제였지만, 빅데이터 시대가 도래하면서 예측분석을 위해 활용되는 PA기법의 중요성이 더욱 부각되었다(Lee et al., 2013).

예측분석을 위해 활용되는 PA기법에는 의사결정나무, 랜덤포레스트, 인공신경망, 회귀분석 등 다양한 기법이 있다. PA기법을 통한 예측분석은 마케팅, 보험, 판매, 의료, 고객센터, IT 등의 다양한 비즈니스에서 널리 사용되고 있으며, 시장 분석, 고객 유지 분석, 리스크 분석, 고객 예측 등을 위해 사용되고 있다(Halper, 2014). PA기법을 활용한 예측분석 관련 연구는 보험의 청구액 예측(Quan & Valdez, 2018), 의료 과실 분석(Frees & Gao, 2019), 소매업의 고객 예측(Bradlow et al., 2017) 등을 예로 들 수 있다.

건설업의 경우에도 PA기법을 이용한 연구가 다수 발견되었으며 공사비 예측(Park & Lee, 2010), 공사 자재 가격 예측(Wang et al., 2019), 생산성 예측(Kim et al., 2011), 건설경

기 예측(Hwang & Kim, 2012) 등이 대표적인 사례이다.

이와 같이 PA기법은 예측이라는 주제를 다루기 위해 유용하게 활용될 수 있는 도구이며, 건축공사에서 발생하는 공중별 사망사고 예측을 위해서도 활용될 수 있다.

2.2.2 의사결정나무

의사결정나무는 PA기법 중 대표적인 기법의 하나로, 의사결정규칙(decision rule)을 나무구조로 도표화하여 특정 목표변수(종속변수)에 대해 관심대상이 되는 모집단을 몇 개의 일원적인 소집단들로 분류하고 예측하기 위해 활용되는 기법이다(Cho et al., 2017). 의사결정나무는 예측분석을 위해 광범위하게 활용되고 있으며, 이는 의사결정나무가 지니고 있는 장점 때문이다. 대표적인 장점으로서는 분석결과가 나무구조로 표현되기 때문에 해당 문제를 직관적으로 이해하고 쉽게 해석할 수 있다는 것이다(Jeong et al., 2015).

의사결정나무를 구축하는 단계와 방법에 대해서는 문헌에 따라 다소 다르게 제시하고 있지만(Song et al., 2015; Cho & Park, 2011) 가장 일반적인 단계와 방법은 다음과 같다.

- 1단계(데이터 수집): 모델 구축에 필요한 데이터를 수집하는 것을 의미한다.
- 2단계(독립변수 및 종속변수 설정): 수집한 데이터를 모델의 목적에 맞게 독립변수와 종속변수로 구분하는 것으로, 연구의 목적에 따라 변수의 성질을 정의하는 작업이다.
- 3단계(기준 설정): 분석 목적과 자료 구조에 따라 알고리즘(algorithm), 최대 나무 깊이 및 최소 케이스 수의 설정 등 모델을 구축하기 위한 기준을 설정하는 작업이다.
- 4단계(가지치기): 분류오류가 커질 위험이 있거나 부적절한 규칙을 가지고 있는 가지를 제거하는 것으로, 오분류율을 낮추기 위해 잘못 분류된 경우를 제거하는 것을 의미한다.
- 5단계(의사결정나무 구축 및 해석): 상기 기준 설정과 가지치기를 바탕으로 의사결정나무 모델을 구축하고 그 결과를 해석하는 것이다.
- 6단계(적합성 평가): 의사결정나무 모델의 적합성을 평가하는 것이다.

본 연구에서 목표로 하고 있는 의사결정나무 기반의 건축공사 사망사고 예측모델의 구축도 상기 과정을 통해 진행되었으며, 주요 내용은 3장 및 4장에서 제시하였다.

2.3 건설 산업재해 관련 선행 연구 분석

건설 산업재해를 주제로 하는 관련 문헌 및 연구를 분석한 결과 다양한 사례들이 발견되었다. 대표적인 예로, Kim

et al. (2017)은 인공지능망 기법을 적용하여 건설공사에서 발생 가능한 건설 산업재해 유형을 예측하였다. Cho et al. (2017)에서는 의사결정나무 기법을 이용하여 건설 산업재해 유형을 예측하는 모델을 구축하였으며, Kim (2020)에서는 랜덤 포레스트 기법을 활용하여 건설 산업재해 유형을 예측하는 모델을 구축하였다.

선행 관련 연구들을 분석한 결과, 대부분 전체적인 산업재해를 연구의 범위로 하였으며, 산업재해 유형별 발생확률을 예측하는 경향을 보이고 있었다. 반면에 건축공사의 어느 공종에서 어느 정도 수준으로 사망사고가 발생할 수 있는지에 대해 예측할 수 있는 예측모델 개발에 관한 연구는 다소 미흡한 것으로 나타났다. 따라서 건축공사의 공중별 사망사고 모델을 구축하고 있는 본 연구에서는 연구범위(건축공사, 사망사고)와 도출되는 결과물(공중별 사망사고 발생확률) 측면에서 선행연구와는 차별성을 지니고 있는 것으로 판단된다.

3. 사망사고 예측모델 구축의 전(前)단계

사망사고 예측모델 구축의 전단계란 2.2.2절에 제시된 6개 단계 중 데이터 수집(1단계)부터 가지치기(4단계)까지를 의미하며 주요 내용은 다음과 같다.

3.1 데이터 수집

의사결정나무를 이용한 건축공사 공중별 사망사고 예측 모델을 구축하기 위해서는 실제 사망사고 데이터가 필요하였다. 본 연구에서는 ‘건설공사 안전관리 종합정보망’을 통해 2019년 7월~2021년 2월의 건축공사 사망사고 사례 124건을 수집하여 모델 구축을 위한 데이터로 활용하였다.

3.2 종속변수 및 독립변수 설정

종속변수란 다른 변수의 영향을 받아 변화 또는 결정되는 변수를 의미하며(Howell, 2010), 본 연구에서는 공중별 사망사고를 예측하는 것을 목표로 하고 있기 때문에 건축공사의 조건에 따른 공중별 사망사고 발생확률을 종속변수로 설정하였다.

이때 건축공사의 공종은 ‘건설공사 안전관리 종합정보망’에 제시된 분류체계를 참조하여 해체 및 철거공사(dismantling and demolition work), 토공사 및 기초공사(earthwork and foundation work), 가설공사(temporary work), 철골공사(steel-frame work), 철근콘크리트공사(RC work), 마감공사(finishing work), 기타공사(others) 등 7개로 분류하였다.

독립변수란 다른 변수에 영향을 받지 않고 독립적으로 변

화하는 변수를 의미한다(Howell, 2010). 본 연구에서 독립변수란 건축공사의 조건을 의미하며 ‘건설공사 안전관리 종합정보망’에 수록된 정보를 참고하여 지역(region), 전체공사비(total construction cost), 해당공종 공사비(construction cost of work), 전체 공사기간(total construction period), 작업프로세스(work process), 공정률(progress rate) 등 6개를 독립변수로 설정하였다. 이때 전체공사비는 해당 공사의 전체공사비를 의미하고 해당공종 공사비는 사망사고가 발생한 공종(예: 철근콘크리트공사)의 공사비를 의미하며, 금액 구간은 두 가지 유형의 금액 범위가 다 포함될 수 있도록 설정하였다.

독립변수의 경우, ‘건설공사 안전관리 종합정보망’에 수록된 정보를 바탕으로 할 경우, 지역, 발생시간, 계절, 요일, 공공/민간, 날씨, 기온, 습도, 시설물 종류, 사고유형, 사고공종, 사고부위, 전체공사비, 해당공사 공사비, 낙찰률, 전체공사기간, 공정율, 작업자수, 안전관리계획 등 19개가 있었다. 따라서 SPSS로 의사결정나무 모델을 초기에 구축할 때는 이들 변수를 모두 포함시켜서 반복적으로 프로그램을 실행하였다.

그러나 초기 결과물들을 확인한 결과, 유효한 독립변수들만 모델에 포함이 되고 나머지 독립변수들은 탈락되는 현상이 발견되었으며, 이는 상대적으로 유효하지 않은 독립변수는 자동 제거됨으로서 모델의 불순도를 낮추고 유효 독립변수만을 포함시키기 때문인 것으로 분석되었다. 이때 불순도란 해당 노드 안에 서로 다른 데이터가 얼마나 섞여 있는지를 나타내는 것으로 불순도가 낮을수록 판별력이 좋다고 해석할 수 있다.

이를 확인하기 위해 의사결정나무 구축 단계 중 3단계(기준 설정)의 기준값을 변경하며 수차례 반복적인 테스트(trial & error)를 진행하였으며, 그 결과 지역, 전체공사비, 해당공종 공사비, 전체 공사기간, 작업프로세스, 공정률 등 6개 독립변수 조합이 가장 유효한 결과를 도출하는 것으로 확인되었다.

상위 과정을 통해 선정된 6개의 독립변수는 세부조건별로 범주화(categorization)하였으며, 이는 <Table 1>에 제시된 바와 같다. 이때 작업프로세스의 경우 기존의 분류된 카테고리 총 24개였으며, 연구 초기단계에서는 작업 프로세스 24개를 모두 사용하여 모델 구축을 시도하였으나 유효한 모델 구축이 어려웠다. 이는 24개 작업 프로세스로 분류하기에는 124개의 데이터가 상대적으로 충분하지 않았기 때문으로 확인되었다. 따라서 가장 빈도수가 높은 6개 카테고리는 그 명칭을 그대로 사용하고 빈도수가 낮은 나머지 18개 카테고리는 기타(others)로 분류하여 모델을 구축하였다.

3.3 기준 설정 및 가지치기

의사결정나무를 활용하여 모델을 구축하기 위해서는 기준 설정과 가지치기가 선행되어야 한다. 기준 설정이란 원하는 모델의 규모를 결정하기 위해 알고리즘(algorithm)과 최대 나무 깊이 및 최소 케이스 수를 설정하는 과정으로, 이에 따라 분리기준과 정지규칙이 결정된다.

분리기준이란 독립변수의 선택과 독립변수의 하위 범주(카테고리)의 병합이 이루어지는 기준이다. 종속변수의 적

Table 1. Independent variables of the model

Variables		Sub-variables by categories	
a	Region	a-1	Seoul
		a-2	Gyeonggi-do and Incheon
		a-3	Gangwon-do
		a-4	Chungcheong-do
		a-5	Jeolla-do
		a-6	Gyeongsang-do
		a-7	Jeju Island
b	Total construction cost	b-1	Less than 100 million won
		b-2	100 million won ~ 1 billion won
		b-3	1 billion won ~ 10 billion won
		b-4	10 billion won ~ 50 billion won
		b-5	More than 50 billion won
		b-6	Others
c	Construction cost of work	c-1	Less than 100 million won
		c-2	100 million won ~ 1 billion won
		c-3	1 billion won ~ 10 billion won
		c-4	10 billion won ~ 50 billion won
		c-5	More than 50 billion won
		c-6	Others
d	Total construction period	d-1	Less than 100 days
		d-2	100~500 days
		d-3	500~1,000 days
		d-4	More than 1,000 days
e	Work process	e-1	Installation work
		e-2	Dismantling work
		e-3	Pouring work
		e-4	Assembly work
		e-5	Painting work
		e-6	Moving
		e-7	Others
f	Progress rate	f-1	0~20%
		f-2	20~40%
		f-3	40~60%
		f-4	60~80%
		f-5	80~100%

정한 분포 및 구별을 위해 필요한 과정이며, 분리기준은 선택하는 알고리즘에 의해 결정된다(Son & Kim, 2012).

정지규칙은 더 이상 분리가 일어나지 않고, 모델이 불필요하게 커지는 것을 방지하기 위한 규칙을 설정하는 것이며(Jeong & Choi, 2014), 정지규칙은 최대 나무 깊이 및 최소 케이스 수에 의해 결정된다(Cho, 2012).

본 연구에서 분리기준은 가능한 한 많은 동질적인 그룹이 속하도록 노드를 구성하는 알고리즘인 CRT (Park et al., 2011)를 선택하였으며, CRT를 통해 종속변수의 분포를 가장 잘 구별해 줄 수 있는 독립변수 구분·분리 기준이 적용되었다. 또한, 최대 나무 깊이와 최소 케이스 수의 경우, 반복적인 테스트(trial & error)를 통해 가장 예측력이 높은 모델 결과를 가져오는 기준을 탐색하여 설정하였다. 이를 통해 예측력이 높으면서 불필요하게 규모가 커지지 않도록 기준을 설정하였다.

의사결정나무 모델을 구축하기 위해서는 가지치기 작업도 필요하다. 가지치기란, 오차가 증가하는 것을 방지하기 위해 모델을 정리하는 작업이다. 가지치기를 수행하지 않을 경우, 오차가 증가할 수 있기 때문이다(Hong et al., 2010). 본 연구에서도 예측모델의 오차를 줄이기 위해 가지치기를 실시할 필요가 있었다. SPSS에서 가지치기의 기능을 활성화 혹은 비활성화 할 수 있으며, 활성화하는 경우 SPSS에서 가지치기 작업이 자동으로 실행 및 중지된다. 본 연구에서는 과하게 학습되어 오차가 증가하는 것을 방지하기 위해 가지치기 기능을 활성화하였다. 이를 통해 오차를 증가시키는 가지를 제거하였다.

상기 과정을 통해 사망사고 예측모델 구축을 위한 1단계~4단계의 전(前)단계 작업이 수행되었으며, 이를 기반으로 건축공사 사망사고 예측모델이 구축되었다.

4. 건축공사 사망사고 예측모델 구축

4.1 모델 구축 결과 및 해석

3장에 제시된 과정 이후, SPSS Statistic 27을 활용하여 의사결정나무 기반의 건축공사 사망사고 예측모델이 <Fig. 1>에 제시된 바와 같이 구축되었다. 의사결정나무는 박스(box) 형태로 표현되는 노드(node)를 기본 구성으로 하며, 각 노드에는 해당 노드의 번호가 부여되고, 노드는 특정 조건에 따라 분류된 분석 결과를 제시하는 역할을 한다.

이때 의사결정나무 모델의 독립변수 분류 우선순위의 경우, SPSS를 통해 자동으로 결정이 된다. 독립변수 분류의 우선순위가 결정되는 기준은 불순도이며, 모델 구축 과정에서 SPSS가 불순도를 내부적으로 계산하여 불순도가 낮은 변수를 상위노드로 배치하게 된다. 즉, 제시된 의사결정나무 모

델은 불순도가 낮은 변수를 상위노드로 배치하는 자동 알고리즘을 통해 일종의 최적화된 결과로 구축된 모델이라고 할 수 있다.

<Fig. 1>에 제시된 모델을 해석하는 방법을 설명하면, 노드 0은 분석 데이터로 사용된 124개 사례를 공사조건에 대한 구분 없이 7개의 공종으로만 구분한 1차 분류이다. 이때, 해당 노드에 속한 전체 데이터(사례) 수를 노드N이라고 하며, 이 경우 노드 0의 노드N은 124개이다.

노드 0의 경우, 사망사고가 가장 많이 발생한 공종은 마감공사(36건, 29.0%)로 나타났다. 이 결과만으로도 마감공사가 가장 빈번하게 사망사고가 발생하는 공종이라는 것을 알 수 있다. 그러나 정교한 분류와 예측을 위해서는 공사조건에 따른 추가 분류가 필요하다.

2차 분류인 노드 1과 노드 2는 공정률(progress rate) 조건으로 124개 데이터가 분류된 결과이며, 공종의 관점에서 유사한 특징을 가진 두 개의 그룹으로 분류되었다. 이때 노드 1의 경우, 82개 데이터가 작업 프로세스(work process) 조건에 따라 하부로 노드 3, 노드 4로 추가적으로 분류되었다. 노드 4의 경우 추가 조건에 따라 하부로 추가 분류가 이루어졌지만, 노드 3의 경우 더 이상 분류 결과가 나타나지 않았다.

노드 3에서 추가 분류가 발생하지 않았던 주된 이유는 노드 3에 사용된 사망사고 사례를 추가 조건으로 분류를 진행하기에는 해당 조건에 부합되는 데이터 수가 적었기 때문이다. 즉, 더 이상 세부적으로 분류하기도 어려울 뿐만 아니라, 자의적으로 분류할 경우, 적은 수의 데이터로 결과를 일반화시킬 오류가 있기 때문이다.

반면, 노드 4에 해당되는 78개 데이터에 대해서는 노드 7~8, 노드 11~12, 노드 15~18로 이어지는 세부 분류가 이루어졌다. 이 경우에도 조건에 부합되는 사망사고 데이터 수에 따라 노드 15~노드 18까지 분류되는 경우도 있지만, 앞에서 설명한 같은 이유로 노드 8처럼 더 이상 세부적인 분류가 일어나지 않은 경우도 발견되었다. 이를 통해 다음과 같은 시사점을 발견할 수 있었다.

- 하부층 노드를 참고하면 보다 정교한 분류와 예측이 가능하다. 예를 들어, <Fig. 1>에 제시된 노드 15의 경우, 노드 1보다 많은 조건(독립변수)에 따른 건축공사 공종별 사망사고 발생확률을 예측할 수 있다.
- 하부층 노드에 부합되는 사례가 부족한 경우라도 상부층 노드를 참고하면 개략적인 예측이 가능하다. 예를 들어, <Fig. 1>에 제시된 노드 1에 분류된 공정률 40% 이상인 경우, 건축공사 공종별 사망사고 발생확률을 개략적으로 가늠할 수 있다.
- 추후 보다 많은 데이터가 축적되면 보다 정교한 분류와 예측이 가능해질 것이며, 그 결과로 보다 정교한 의사결

정나무 모델의 구축이 가능해진다. 즉, 모델의 지속적인 업데이트가 필요한 것이다.

(Fig. 1)에 제시된 건축공사 사망사고 의사결정나무 모델을 예측이라는 관점에서 활용하는 방법은 반응(response)과 지수(index)라는 두 가지 방법이 있으며 각각에 대한 설명은 아래와 같다.

4.2 반응을 통한 예측 및 활용

반응(response)을 통한 예측이란 의사결정나무에 표현된 반응 값을 활용하여 예측을 하는 방법이다. 이때 반응이란 특정 노드에서 종속변수의 발생확률(%)을 의미하며 의사결정나무에서 '%'로 표현된다. 예를 들어 노드 15의 경우, 가설공사(temporary work)에서 사망사고가 발생할 확률은 28.6%이다. 반응을 통한 예측은 특정 공사조건을 가정하고, 공종별 사망사고의 발생확률을 예측하는데 유용하다.

예를 들어, '전체공사비(total construction cost) 425억원, 해당 공종공사비(construction cost of work) 9억원, 전체공사기간(total construction period) 526일인 현장의 공정률(progress rate) 50%이면서 운반작업(moving)을 하고 있는 경우, 사망사고가 발생확률이 가장 높은 공종은 무엇이며, 각 공종별 사망사고 발생확률은 얼마일까?'라는 질문을 모델의 상부층에서 하부층로 내려가면서 설명하면 다음과 같다.

- 모든 예측의 출발은 노드 0에서 시작한다.
- 공정률 50%는 노드 1에 해당된다.
- 작업프로세스가 운반작업(moving)인 경우는 노드 4에 해당된다.
- 전체공사기간 526일은 노드 7에 해당된다.
- 전체공사비 425억원은 노드 11에 해당된다.
- 해당공종 공사비 9억은 노드 15에 해당된다.

상기 예시를 보면 질문의 조건에 해당되는 노드는 노드 15로 나타난다. 이때 노드 박스 안에는 위에서 설명한 반응(%)와 함께 'n'으로 표시되는 값이 있으며 이를 이익N(gain N)이라고 한다. 이익N이란 수집된 데이터 중 해당 공사조건 및 공종에서 사망사고가 발생한 데이터(사례) 수이고, 반응은 사망사고 공종별 발생가능성(확률)을 의미하며 '(이익N/노드N)×100(%)'로 산정된다.

상기에 질문 형식으로 제시된 공사조건에 해당되는 노드 15의 경우를 살펴보면, 사망사고가 발생할 가능성이 가장 높은 공종은 가설공사(temporary work), 해체 및 철거공사(dismantling and demolition work), 기타공사(others) 모두 동일하게 28.6%(반응)로 나타난다. 즉, 3개 공종에서 사망사고가 발생할 확률이 동일하게 28.6%라는 것을 의미한다. 이 조건에 해당되는 현장이라면, 이들 공종에서 사망사고 발생

Table 2. Responses of terminal nodes

(Unit : %)

Type of construction works	Node 3,13,20	Node 5	Node 9	Node 15	Node 17	Node 18	Node 19
RC work	100.0	0.0	28.6	7.1	0.0	16.7	21.1
Temporary work	0.0	0.0	71.4	28.6	0.0	11.1	5.3
Steel-frame work	0.0	0.0	0.0	0.0	7.7	44.4	10.5
Dismantling and demolition work	0.0	100.0	0.0	28.6	0.0	0.0	10.5
Earthwork and foundation work	0.0	0.0	0.0	0.0	7.7	5.6	47.4
Finishing work	0.0	0.0	0.0	7.1	76.9	22.2	5.3
Others	0.0	0.0	0.0	28.6	7.7	0.0	0.0

Table 3. Independent variables of terminal nodes

Node	Independent variables
3	<ul style="list-style-type: none"> ■ Progress rate : 40~60%; 60~80%; 80~100% ■ Work process : pouring work;
5	<ul style="list-style-type: none"> ■ Progress rate : 0~20%; 20~40% ■ Work process : dismantling work
9	<ul style="list-style-type: none"> ■ Progress rate : 0~20%; 20~40% ■ Work process : installation work
13	<ul style="list-style-type: none"> ■ Progress rate : 0~20%; 20~40% ■ Work process : pouring work; carrying work
15	<ul style="list-style-type: none"> ■ Progress rate : 40~60%; 60~80%; 80~100% ■ Work process : installation work; cleanup work; dismantling work; others; moving; assembly work; carrying work; painting work ■ Total construction period : 100 ~ 500 days; 500 ~ 1,000 days ■ Total construction cost : 10 billion won ~ 50 billion won; less than 100 million won; more than 50 billion won ■ Construction cost of work : 100 million won ~ 1 billion won; less than 100 million won
17	<ul style="list-style-type: none"> ■ Progress rate : 40~60%; 60~80%; 80~100% ■ Work process : installation work; cleanup work; dismantling work; others; moving; assembly work; carrying work; painting work ■ Total construction period : 100 ~ 500 days; 500 ~ 1,000 days ■ Total construction cost : 1 billion won ~ 10 billion won; 100 million won ~ 1 billion won; others ■ Region : Gangwon-don; Seoul; Chungcheong-do
18	<ul style="list-style-type: none"> ■ Progress rate : 40~60%; 60~80%; 80~100% ■ Work process : installation work; cleanup work; dismantling work; others; moving; assembly work; carrying work; painting work ■ Total construction period : 100 ~ 500 days; 500 ~ 1,000 days ■ Total construction cost : 1 billion won ~ 10 billion won; 100 million won ~ 1 billion won; others ■ Region : Gyeongsang-do; Gyeonggi-do and Incheon; Jeju Island
19	<ul style="list-style-type: none"> ■ Progress rate : 0~20%; 20~40% ■ Work process : cleanup work; others; moving; assembly work; painting work ■ Construction cost of work : 100 million won~ 1 billion won; 1 billion won~ 10 billion won; less than 100 million won; others
20	<ul style="list-style-type: none"> ■ Progress rate : 0~20%; 20~40% ■ Work process : cleanup work; others; moving; assembly work; painting work ■ Construction cost of work : 10 billion won ~ 50 billion won; more than 50 billion won

확률을 가장 염두에 두고 현장을 관리할 필요가 있다는 의미이다.

이상에서 반응 값을 이용하여 공종별 사망사고 가능성을 예측하는 예시를 제시하였으며, 이를 <Fig. 1>에 제시된 결과를 중심으로 설명하기 위해 최종 노드(terminal node)의 반응 값을 정리하여 제시하면 <Table 2>와 같으며 <Table 3>은 각 최종 노드의 공사조건이다.

<Table 2>를 통해 나타나는 주요 특징과 시사점을 요약하면, 공종별 사망사고 발생확률은 공사조건에 따라 다르게 발생하며, 따라서 각 공사조건에 맞는 관리가 필요하다는 것이다. 또한, 특정 공종에 사망사고 발생확률이 집중될 경우(예: 노드 5) 특정 공종을 집중 관리해야 할 필요가 있으며, 다수 공종의 사망사고 발생확률이 비슷한 경우에는(예: 노드 15) 다수의 공종을 전반적으로 관리할 필요가 있다. 즉, 건축공사에서 사망사고의 선제적인 관리를 위해 어떤 공종에서 선제적으로 관리하고 대비해야 하는지를 예측모델을 통해 파악할 수 있다.

4.3 지수를 통한 예측 및 활용

의사결정나무 기반의 모델로 공종별 사망사고를 예측하는 두 번째 방법은 지수(index) 값을 활용하는 것이다. 지수는 이익도표에 표현되며, 이익도표란 각 노드별 분석결과 정보를 제시하는 도표로, 해당 건축공사 사망사고의 분석결과를 노드N, 이익N, 노드퍼센트(node percent), 이익퍼센트(gain percent), 반응(response), 지수(index) 등의 6가지 수치로 표현된다. 이때 이익도표는 의사결정나무 모델에는 표시되지 않고 별도의 결과물로 산출된다.

지수가 의미하는 바는 특정 노드에서의 종속변수별 발생확률이 전체 종속변수별 발생확률보다 몇 배가 높은지를 의미하며 이익도표에서 ‘%’로 표현된다.

지수를 산정하는 공식은 다양한 형태로 표현될 수 있지만 가장 직관적인 공식은 ‘특정 노드의 반응(response)/노드 0의 반응(response) × 100’의 비율이며, 특정 노드의 특정 종속변수 발생확률이 전체(노드 0)의 특정 종속변수 발생확률에 비해 몇 배 높은지에 대한 정보를 나타낸다. 본 연구 맥락에서는 특정 공사조건의 특정 공종에서 사망사고 발생확률이 특정 공종 전체에서 사망사고 발생확률에 비해 높은 정도를 의미한다.

예를 들어 노드 15 가설공사(temporary work)의 경우, 사망사고가 발생할 확률은 전체 가설공사 사망사고 발생확률보다 236.4%(약 2.36배) 높으며 이는 다음과 같은 과정을 통해 산출된 것이다.

가설공사의 경우, 노드 15에서 반응은 28.6%이며, 노드 0의 반응은 12.1%이므로, 해당 경우의 지수는 ‘(28.6/12.1)×

100(%)’로 약 236.4%가 된다. 이는 노드 15 조건에 해당되는 현장의 가설공사에서 사망사고가 발생할 확률이, 전체 가설공사에서 사망사고가 발생할 확률보다 약 2.4배(236.4%) 정도 높음을 의미한다. 다른 조건의 현장에 비해 노드 15에 해당되는 현장의 가설공사에서 사망사고를 피하기 위한 주의와 노력이 더 필요하다는 것을 의미한다.

상기에 제시된 바와 같이 지수를 이용한 예측은 특정 공종을 가정하여, 노드별 사망사고 발생확률이 전체와 비교하여 몇 배가 높은지를 예측하는데 유용하다.

이상에서 지수를 이용하여 공종별 사망사고 가능성을 예측하는 예시를 제시하였으며, 각 최종 노드의 이익도표에 제시된 지수 값을 정리하여 제시하면 <Table 4>와 같으며, 각 최종 노드의 공사조건은 이미 제시한 <Table 3>과 같다.

Table 4. Index of terminal nodes

(Unit : %)

Type of construction works	Node 3,13,20	Node 5	Node 9	Node 15	Node 17	Node 18	Node 19
RC work	413.3	0.0	118.1	29.5	0.0	68.9	87.0
Temporary work	0.0	0.0	590.5	236.4	0.0	91.9	43.5
Steel-frame work	0.0	0.0	0.0	0.0	79.5	459.3	108.8
Dismantling and demolition work	0.0	1127.3	0.0	322.1	0.0	0.0	118.7
Earthwork and foundation work	0.0	0.0	0.0	0.0	79.5	57.4	489.5
Finishing work	0.0	0.0	0.0	24.6	265.0	76.5	18.1
Others	0.0	0.0	0.0	442.9	119.2	0.0	0.0

<Table 3>을 통해 지수의 공종별 사망사고 발생확률을 예측하는 방법은 가로축을 따라 해석하면 된다. 예를 들어, ‘가설공사 전체의 사망사고 발생확률과 비교했을 때, 어떤 조건의 건축공사에서 가설공사 사망사고의 발생확률이 높은가?’라는 예시를 <Table 4>의 가로축을 따라 찾아보면 노드 9에 해당된다.

이는 노드 9의 경우, 가설공사 전체보다 사망사고 발생확률이 약 5.91(590.5%)배 높다는 것을 의미하며, 그다음 노드 15(2.36배), 노드 18(0.92배), 노드 19(0.44배) 순으로 나타난다. 이때 노드 18과 노드 19와 같이 지수 값이 100% 미만인 경우 선제적 관리 및 대응의 필요성이 노드 9, 노드 15에 비해 낮음을 의미하며, 지수 값이 클수록 특정 공종(종속변수) 전체의 발생확률보다 더 높은 수치이기 때문에 이 경우 선제적 관리 및 대응의 필요성이 높아진다는 것을 의미한다. 또한, 지수의 경우 특정 공종 전체의 사망사고 발생확률과 비교하여, 사망사고 발생확률이 높은 공사조건을 예측하여 대비할 수 있기 때문에, 전사적(본사) 측면에서 공종별 사망사고를 집중적으로 관리해야 할 공사를 파악할 때 활용

가능하다.

(Table 4)를 통해 나타나는 특징과 시사점을 요약하면, 특정 공종에서, 특정 현장이 전체 사망사고 발생확률보다 얼마만큼 높은 발생확률을 지녔는지는 공사조건에 따라 다르다는 것이다. 즉, 공사조건에 따라 전체 사망사고 발생확률보다 높은 발생확률을 지니는 경우도 있으며, 반대로 전체 사망사고 발생확률보다 낮은 발생확률을 지닌 경우도 있다는 것이다. 이는 전사적(본사) 관점의 선제적 관리를 위해 현장들의 사망사고 관리 우선순위를 도출하고자 할 때, 지수 값이 활용될 수 있음을 의미한다.

상기에 제시된 바와 같이 반응을 통한 예측과 지수를 통한 예측은 각기 다른 관점에서 활용될 수 있다. 반응의 경우, 현장조건에 따른 공중별 발생확률을 파악하기에 적절하며, 지수의 경우, 사망사고의 발생확률이 높은 공사의 조건과 유형을 파악하는데 유용하다. 따라서 사망사고의 발생확률을 예측하기 위해 반응과 지수를 병행하여 사용하는 것이 바람직하다고 할 수 있다.

4.4 모델 적합성 평가

모델 적합성 평가란 의사결정나무 모델의 적합성을 평가하는 것이다. 적합성 평가에는 다양한 방법들이 있지만 본 연구에서는 수치적으로 적합성을 평가할 수 있는 위험추정치(risk estimate) 방법을 채택하였다. 주된 이유는 정량적으로 적합성을 판단할 수 있다는 장점으로 인해 널리 채택되고 있는 방법이기 때문이다(Choi et al., 2011; Koo & Kim, 2015; Jang, 2014).

위험추정치란, 의사결정나무에 의해 잘못 분류되거나 잘못 예측될 위험수준을 수치적으로 표현한 것이며, 이 값이 작을수록 적합하다는 것을 의미한다. 예를 들어, 위험추정치 0.40은 분류된 데이터 중 40%는 잘못 분류되었으며, 분류된 나머지 데이터 60%는 적합하게 분류되었음을 의미한다.

이때 적합성을 판단하는 절대적인 기준선은 없는 것으로 조사되었다. 그러나 다수의 선행연구들을 참조한 결과, 일반적으로 0.40 이하의 위험추정치를 보인다면, 해당 모델은 적합한 것으로 나타나고 있다(Kweon & Kim, 2014; Lee et al., 2012; Oh & Oh, 2018). 본 연구의 건축공사 사망사고 의사결정나무 모델의 경우, 위험추정치가 0.379로 SPSS를 통해 산출되어 일정 수준의 적합성을 확보한 것으로 평가되었으며, 추후 데이터 수의 증가에 따른 지속적인 모델 업데이트를 통해 예측도가 향상될 수 있을 것으로 기대된다.

4.5 요약 및 시사점

의사결정나무 기반의 사망사고 예측모델 구축과 반응 및 지수를 통한 예측에서 나타나는 주요 특징과 시사점을 요약

하면 다음과 같다.

- 의사결정나무를 활용하여 건축공사 공중별 사망사고를 예측할 수 있는 모델의 구축 가능성을 확인할 수 있었다. 이는 사망사고 예측뿐만 아니라 건설과 관련된 다양한 예측이라는 주제에 의사결정나무가 활용될 수 있다는 것을 시사한다.
- 건축공사의 조건에 따라 공중별 사망사고의 발생확률이 다르다는 것을 파악할 수 있었으며, 이는 각 공사조건에 따라 적합한 선제적인 관리가 필요하다는 것을 의미한다.
- 건축공사의 조건에 따라 사망사고를 줄이기 위해 특정 공중에 집중하여 관리해야 하는 경우와 다수의 공종을 전반적으로 관리해야 하는 경우가 있다는 것을 확인할 수 있었다.
- 공사조건(독립변수)이 증가할수록 정교한 예측이 가능하고, 이에 준하는 데이터 수가 필요하다는 것을 알 수 있었으며, 이를 위해 향후 지속적인 신규 데이터 축적 및 업데이트의 필요성을 확인할 수 있었다.
- 건축공사 공중별 사망사고 예측을 위해 반응과 지수를 활용하는 방법을 검토하고 각각 현장(반응)과 본사(지수)에서 사용될 수 있다는 것을 확인하였으며, 이는 예측 모델을 효과적으로 사용하기 위해서 두 가지 방법이 보완적으로 활용될 수 있다는 것을 의미한다.

5. 결론

건설현장에서 사망사고를 줄이는 것은 건설사에게는 매우 중요한 현안이며 이를 위해서는 선제적인 예측과 대응이 필요하다. 이를 위해 본 연구에서는 건축공사 공중별 사망사고 예측모델을 구축하여 제시하였다.

건축공사 공중별 사망사고 예측모델의 개발을 위해 실제 사망사고 사례 데이터를 활용하였으며, PA기법 중 의사결정나무를 활용하였다. 구축된 의사결정나무 기반의 건축공사 공중별 사망사고 예측모델을 해석하는 방법과 반응 및 지수를 통한 예측 방법도 제시하였다.

본 연구를 통해 의사결정나무 기반의 예측모델이 건축공사 공중별 사망사고 예측을 위해 어떻게 활용될 수 있는지 제시하였고, 의사결정나무가 건설 관련 다양한 예측에 활용될 수 있는 가능성도 확인할 수 있었다. 또한, 예측모델의 향상을 위해 향후 데이터 축적을 통한 지속적인 모델 업데이트의 필요성도 제안하였다.

건축공사 공중별 사망사고 예측모델은 건축현장과 본사에서 사망사고가 발생할 수 있는 공종과 발생확률을 예측하고 이에 적합한 선제적 대응조치 마련을 위해 유용하게 활

용될 수 있으며 이를 통해 건축공사 사망사고 감소에 기여할 수 있을 것으로 기대된다.

감사의 글

이 성과는 정부(과학기술정보통신부)의 재원으로 한국연구재단의 지원을 받아 수행된 연구임 (No. 2020R1F1A1068593).

References

- Bae, S.Y., and Lee, C. (2018). "A Case Study on the Major Accident in Construction Projects." *Proceedings of the National University Student Conference*, KICEM, 1(13), pp. 48-51.
- Baek, S.J. (2017). "An Predictive Analytics based on Goal-Scenario for Self-adaptive System." *Journal of the Korea Convergence Society*, Korea Convergence Society, 8(11), pp. 77-83.
- Bradlow, E.T., Gangwar, M., Kopalle, P., and Voleti, S. (2017). "The Role of Big Data and Predictive Analytics in Retailing." *Journal of Retailing*, Elsevier, 93(1), pp. 79-95.
- Cho, J.S. (2012). "Inflow and Outflow Analysis of Double Majors using Social Network Analysis." *Journal of the Korean Data & Information Science Society*, Korean Data & Information Science Society, 23(4), pp. 693-701.
- Cho, K.H., and Park, H.C. (2011). "A Study on Decision Tree Creation using Intervening Variable." *Journal of Korean Data Information Science Society*, Korean Data Information Science Society, 22(4), pp. 671-678.
- Cho, Y., Kim, Y.C., and Shin, Y. (2017). "Prediction Model of Construction Safety Accidents using Decision Tree Technique." *Journal of the Korea Institute of Building Construction*, Korean Institute of Building Construction, 17(3), pp. 295-303.
- Choi, S.I., Park, T.J., and Kang, J.G. (2011). "A Study on the Housing Types Selection Expectation in Senescence using a Decision Tree - Focus on the Baby Boom Generation in Pusan." *Journal of The Residential Environment Institute of Korea*, Residential Environment Institute Of Korea, 9(2), pp. 235-251.
- Frees, E.W., and Gao, L. (2019). "Predictive Analytics and Medical Malpractice." *North American Actuarial Journal* 24(2), Society of Actuaries pp. 211-227.
- Halper, F. (2014). *Predictive Analytics for Business Advantage*, TDWI Best Practices Report.
- Hong, A., Ko, J., Yoo, S. (2010). "A Study on the Forecasting Model of the Investment Characteristics of Seoul Office Buildings based on Data Mining." *Seoul Studies*, Seoul Institute, 11(2), pp. 51-68.
- Howell, D.C. (2010). *Fundamental Statistics for Behavioral Sciences*, 8th ed, Cengage Learning.
- Hwang, N.H., and Kim, Y.S. (2012). "Modeling and Prediction of Construction Business using Data Mining Technique." *Proceedings of the National University Student Conference*, KICEM, pp. 40-43.
- Jang, H.D. (2014). "Determinants of Intention to Moving and Remodeling in Small and Medium-Sized Apartments using Decision-tree Analysis." *Journal of the Architectural Institute of Korea (JAIK)*, Architectural Institute Of Korea, 30(9), pp. 45-56.
- Jeong, C, Jeong, W.Y., and Shin, D. (2015). "Selection of the Optimal Decision Tree Model using Grid Search Method : Focusing on the Analysis of the Factors Affecting Job Satisfaction of Workplace Reserve Force Commanders." *Journal of the Korean Operations Research and Management Science Society*, Korean Operations Research and Management Science Society (KORMS), 40(2), pp. 19-29.
- Jeong, Y.H., and Choi, Y.W. (2014). "A Study on the Analysis of Urban Highways Traffic Accident's Impact Factors based on Building Discriminant Models-In Busan Metropolitan City." *Journal of The Korean Society of Civil Engineers*, Korean Society of Civil Engineers, 34(4), pp. 1269-1278.
- Kim, E.J. (2020). "Prediction Model for Construction Safety Accidents using Random Forest." *Journal of the Regional Association of Architectural Institute of Korea*, Regional Association of Architectural Institute of Korea, 22(5), pp. 295-303.
- Kim, H.M., Kim, T.H., Shin, Y.K., Kim, Y.S., and Han, S.W. (2011). "Regression Technique-based Productivity Estimation Conducting Construction Delay Factor Analysis on Interior Works in High-rise Building Construction." *Proceedings of the Korean Institute of Building Construction Conference*, Korea Institute of Building Construction, 11(1), pp. 323-324.
- Kim, Y.C., Yoo, W.S., and Shin, Y. (2017). "Application of Artificial Neural Networks to Prediction of Construction Safety Accidents." *Journal of the Korean Society of Hazard Mitigation*, Korean Society of Hazard Mitigation, 17(1), pp. 7-14.
- Koo, K.M., and Kim, C.J. (2015). "Analysis on Participation Factors for Physical Activity of People with Brain Lesion by using Decision Tree Analysis." *Journal of Sport and Leisure Studies*, Korean Society of Sport and Leisure Studies, 60, pp. 633-643.
- Kweon, Y.R., and Kim, S.Y. (2014). "Predictors of Protective Factors for Internet Game Addiction in Middle School

- Students using Data Mining Decision Tree Analysis.” *Journal of Korean Academy of Psychiatric and Mental Health Nursing*, Korean Academy of Psychiatric and Mental Health Nursing, 23(1), pp. 12-20.
- Lee, C.H., Hur, J., Oh, H.J., Kim, H.J., Ryu, P.M., and Kim, H.K. (2013). “Technology Trends of Issue Detection and Predictive Analysis on Social Big Data.” *Electronics and Telecommunications Trends*, Electronics and Telecommunications Research Institute, 28(1), pp. 62-71.
- Lee, C.S., Jung, M.N., and Kim, Y.J. (2012). “Predictors of Suicidal Ideation for Korean Middle and High School Students : The Application of Decision Tree Analysis.” *Studies on Korean Youth*, National Youth Policy Institute, 23(1), pp. 31-55.
- Ministry of Employment and Labor (MEL) (2021). *Status of Industrial Accidents at the end of December 2020*.
- Mishra, N., and Silakari, S. (2012). “Predictive Analytics: A Survey, Trends, Applications, Opportunities & Challenges.” *International Journal of Computer Science and Information Technologies*, AIRCC, 3(3) pp. 4434-4438.
- Oh, J.A., and Oh, H. (2018). “The Predictors of Factors Related to Career Decision Making Amongst Adolescents who drop out of School, using Decision-Making Tree Analysis.” *Studies on Korean Youth*, National Youth Policy Institute, 29(1), pp. 145-177.
- Park, J.H., and Lee, H.K. (2010). “The Accuracy Analysis of Data Mining Cost Prediction Methods by Cost Factors Classification.” *Journal of the Regional Association of Architectural Institute of Korea*, Regional Association of Architectural Institute of Korea, 12(3), pp. 301-308.
- Park, S.H., Kim, S.S., and Hwang H.S. (2011). *Understanding and Utilization of Advanced SPSS*, Hannarae.
- Quan, Z., and Valdez, E.A. (2018). “Predictive Analytics of Insurance Claims using Multivariate Decision Trees.” *Dependence Modeling*, De Gruyter Open Ltd. 6(1), pp. 377-407.
- Sa, Y.B., Choi, S.U., Cho, W.C., and Lee, T.S. (2012). “Cost Analysis of Fall Accidents in Domestic Construction Industry.” *Journal of Korean Society of Societal Security*, Korean Society of Disaster & Security, 5(1), pp. 1-6.
- Son, Y., and Kim, H. (2012). “Forecasting Export & Import Container Cargoes using a Decision Tree Analysis.” *Journal of Korea Port Economic Association*, Korea Port Economic Association, 28(4), pp. 193-207.
- Song, Y.Y., Lu, Y. (2015). “Decision Tree Methods: Applications for Classification and Prediction.” *Shanghai Archives of Psychiatry*, Editorial Board of Shanghai Archives of Psychiatry, 27(2), pp. 130-135.
- Wang, C., Xinyi, Z., Minggang, W., Ming, K.L., and Pezhman, G. (2019). “Predictive Analytics of the Copper Spot Price by Utilizing Complex Network and Artificial Neural Network Techniques.” *Resources Policy*, Elsevier, 63, 101414, pp. 1-17.
- Yang, Y.K., and Kim, B.S. (2014). “An Analysis of Influential Factors and their Prioritization in Association with the Loss from Construction Disasters with a Focus on Uninsured Categories.” *Journal of the Korea Safety Management & Science*, Korea Safety Management & Science, 16(3), pp. 23-34.

요약 : 건설업은 타 산업에 비해 높은 사망사고율을 보이고 있으며, 최근 사망사고에 대한 기업의 법적 책임이 강화되고 있다. 이는 건설사에게 큰 부담이 되고 있으며, 건설 사망사고에 대한 선제적 예측과 관리의 필요성을 증대시키고 있다. 본 연구의 목적은 의사결정나무를 활용하여 건축공사에서 발생할 수 있는 사망사고를 사전에 예측할 수 있는 모델을 개발하는데 있다. 본 연구에서 의미하는 사망사고 예측모델이란 건축공사의 조건에 따라 공중별로 발생할 수 있는 사망사고의 가능성(확률)을 예측하는 모델을 의미한다. 예측모델의 활용을 통한 사망사고에 대한 사전 예측과 선제적 대응은 법적 처벌을 예방하고 건축공사의 성공적인 수행이라는 측면에서 중요한 의의를 지닌다.

키워드 : 예측분석, 의사결정나무 모델, 건설 산업재해, 건설 사망사고
