

# 산업재해 감소를 위한 안전사고 예측



김종현 김한열 박지환 이진규

<b>1. 연구소개</b>	_____	1-1. 연구 배경 1-2. 현황 1-3. 심각성
<b>2. 연구방법</b>	_____	2-1. 분석 순서도 2-2. 문헌조사 2-3. 변수 확인 2-4. 데이터 분석 – 기초통계량 2-5. 데이터 전처리
<b>3. Modeling</b>	_____	3-1. Machine Learning 3-2. Deep Learning
<b>4. 결론</b>	_____	4-1. 연구 결과 및 시사점 4-2. 정책 제언 및 기대효과 4-3. 연구의 한계점 및 향후 연구 제언
<b>5. 참고문헌</b>	_____	5-1. 참고 문헌

## 연구소개

### 1-1 연구 배경

건설 분야에서의 중대재해는 매년 끊이지 않는 사고로, 긴 공사기간동안 익숙해지는 부주의와 과도한 재하청 등의 복합적인 원인으로 발생하고 있다.

비교적 최근 건설업 분야에서 산업재해의 심각성을 보여주는 사건으로 2022년 발생한 광주 화정 아이파크 붕괴 사고(2022.1.11)가 있다. 날씨를 고려하지 않은 무리한 공사 일정과 비용절감을 이유로 여러 번의 재하청을 거치며 부실한 설계를 용인하였고, 다수의 사상 사고로 사회에 큰 파장을 불러일으켰다. 해당 사고가 발생하기 7개월 전에는 광주에서 철거작업을 하던 건물이 붕괴되는 사고가 있었으며, 2023년에는 광주 아이파크 붕괴 사고에서 화두가 되었던 동일한 무량판 구조로 설계된 검단 신도시의 아파트 붕괴사고가 있었다.



1) 2022년 광주 화정 아이파크 붕괴 사고



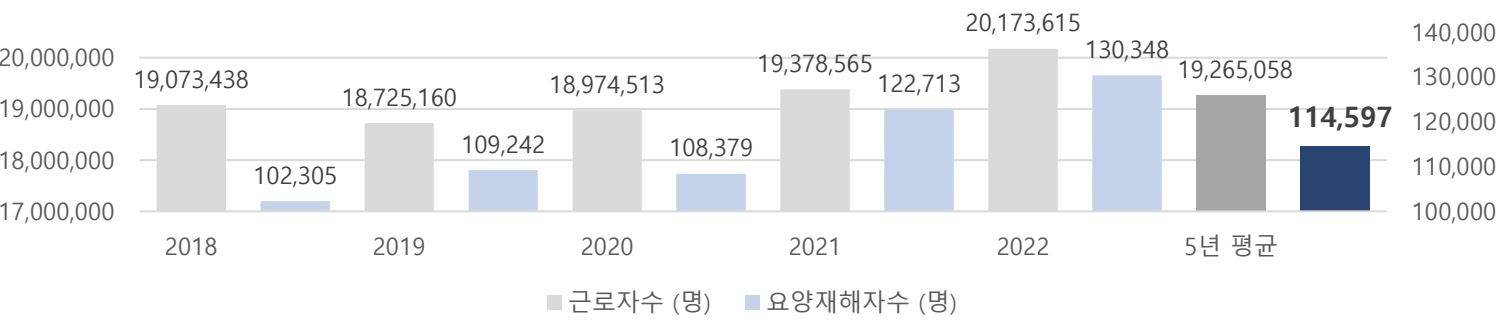
2) 2021년 광주 철거 건물붕괴 사고

고용노동부의 「산업재해 현황분석」에 따르면 산업재해 발생 현황은 1988년부터 2003년까지 증가 추세를 보이다가 2003년부터 2017년까지 꾸준히 감소 추세를 보였으나 2017년 이후 다시 증가 추세를 보인다(e-나라지표 n.d.).

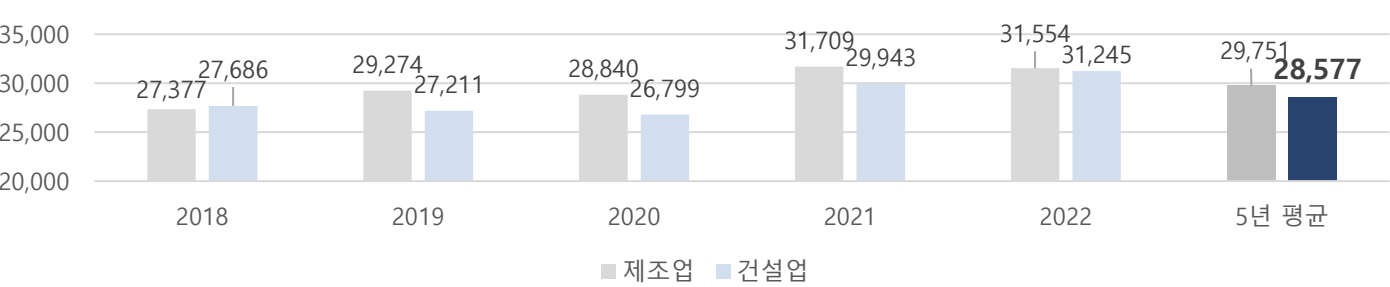
전체 산업군에서 근로자 수는 전년대비 1~4%의 증가율을 보이는 반면, <sup>1)</sup>요양재해자 수는 6~13%의 증가율을 보이며 한번의 사고로 인적, 물적 피해가 큰 제조업과 건설업 모두 증가 추세를 보이고 있다. 5년 평균 요양재해자 수는 제조업 29,751명(25.96%), 건설업 28,577명(24.94%)으로 전체 요양재해자 수(114,597명)의 약 51%을 차지하고 있다. (KOSIS n.d.)

<sup>1)</sup> 요양재해자 수: 근로복지공단에서 요양승인된 재해자수(지방고용노동관서 산재미보고 적발재해 포함)를 말함 (e-나라지표. n.d)  
 \* 요양재해자 수 = 사고재해자 수 + 질병재해자 수

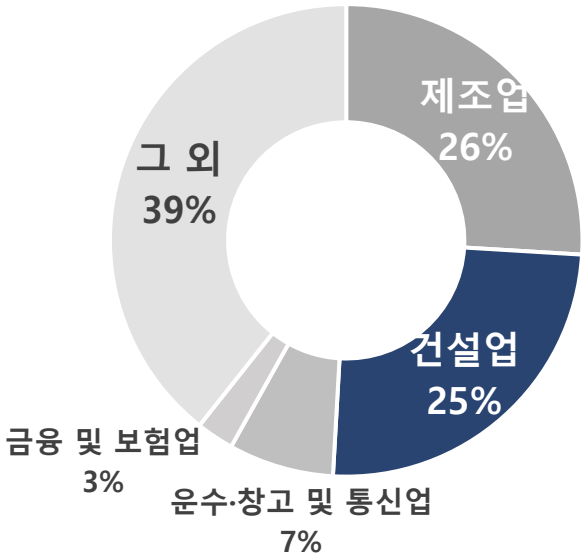
(‘18~’22) 전체 근로자 수, 전체 요양재해자 수 (명)



(‘18~’22) 제조업, 건설업 요양재해자 수 (명)



(‘18~’22) 5년 평균 요양재해자수 (명)



# 연구소개

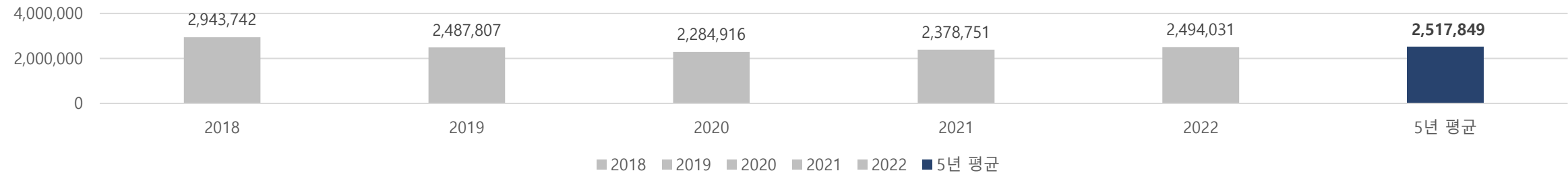
## 1-3 심각성

건설업의 근로자 수는 2018년 이후로 감소하는 추세를 보이고 있다. 반면에 요양재해율을 전반적인 증가 추세를 보이는 것으로 보아 현장에서 아직도 많은 사고가 발생하는 것으로 예상된다.

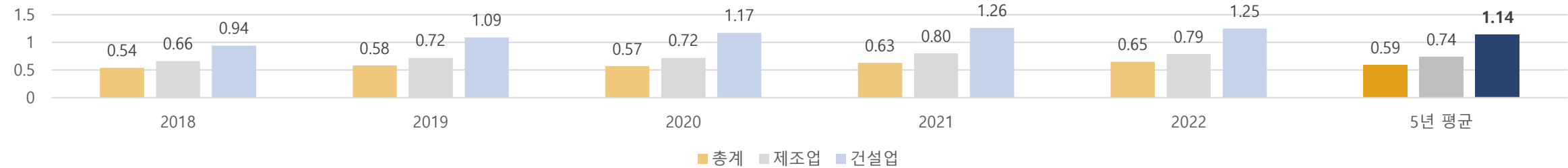
(\*요양재해율의 경우, 5년 평균 0.59%에 비해 제조업 0.74%, 건설업 1.14%로 약 1.5~2배로 높은 수치를 보인다. (KOSIS n.d.))

\*요양재해율 = 요양재해자 수 / 근로자 수 \* 100

건설업 근로자 수 (명)



전체 산업, 제조업, 건설업 요양재해율 (%)



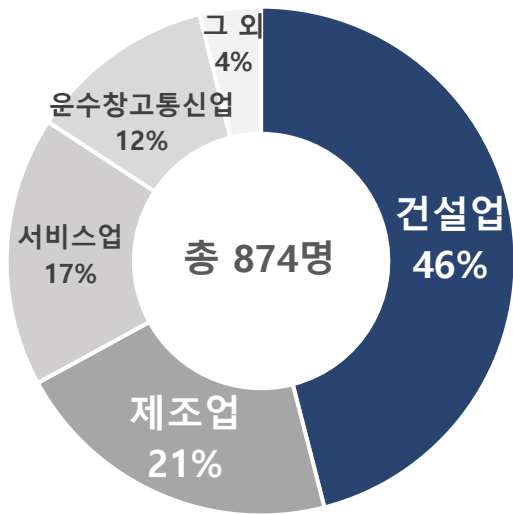
# 연구소개

## 1-3 심각성

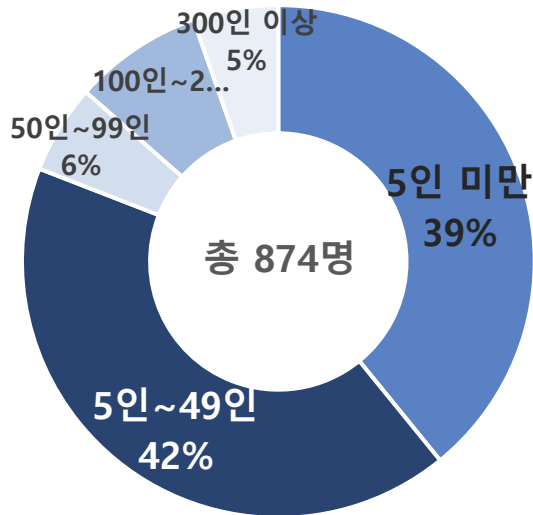
근로자의 안전을 보장하기 위하여 2022년 중대재해처벌법이 시행되었지만 안전보건 관리공단의 「통계로\_보는\_2022년\_산업재해」에 따르면 2022년 전체 업종에서 발생한 사고사망자 874명 중 약 절반 가량인 402명이 건설업에서 발생하였고 두 번째로 높은 184명이 제조업에서 발생하였다. 중대재해처벌법의 기준인 50인 이상 사업장에서 사고사망자는 총 874명 중 167명(19.11%)이 발생한 반면 해당하지 않는 50인 미만 사업장에서 총 874명 중 707명(80.89%)이 발생하였다.

산업재해가 발생한 공사현장의 데이터와 날씨 데이터를 활용하여 사고 발생 위험이 높은 인적사고 유형 예측하여 사고의 예방과 재발 방지에 경각심을 인지시키고자 본 연구를 진행하였다.

전체 업종별 사고사망자 수(명)



전체 업종내 규모별 사고사망자 수(명)





크롤링, API 등의 경로를  
이용한 데이터 수집



EDA를 통한 수집한 데이터의  
1차 분석과 분석 결과를  
통한 전처리 방법 결정 및  
데이터 전처리 진행



전처리를 진행한 데이터에  
머신러닝 모델 훈련



모델 성능 평가 및  
최종 결론 도출



	건설재해 예측 모델 개발	
논문	건설현장 정형·비정형 데이터를 활용한 기계학습 기반의 건설재해 예측 모델 개발 (조민건, 이동환, 박주영, 박승희)	의사결정나무기법을 이용한 건설재해 사전 예측모델 개발 (조예림, 김연철, 신윤석)
데이터	건설공사 안전관리 종합정보망 사고사례 (6,826건, 국토교통부)	2014년 발생한 건설 현장 재해 사례(13,232건)
종속변수	6가지 재해유형 (떨어짐, 넘어짐, 물체에 맞음, 끼임, 절단 및 베임, 부딪힘)	6가지 재해유형 (떨어짐, 넘어짐, 맞음, 끼임, 절단, 부딪힘)
독립변수	날씨(기온, 습도, 계절), 시간, 공사특성(공사종류, 공종,시설, 작업장소, 프로세스, 사고원인)	계절, 요일, 시간, 지역, 작업자특성(성별, 나이, 근무기간, 국적, 고용형태,직무) 공사특성(공사종류, 규모, 공정률, 근로자수)
결론 (Accuracy)	79.98 %(정형 독립변수, DF) 95.41 %(비정형 독립변수 '사고원인' 추가,DF) 사용분류모델(Logit, SVM, NB, ANN, DF)	80%(의사결정모델) 63%(판별분석)
한계점	데이터 불균형 문제, 데이터 표본 부족, 과적합 여부 미제시	사용 모델 부족, 데이터 표본 출처 부정확, 과적합 여부 미제시



연구방법

2-2

문헌 조사 – 위험도 평가 지수 개발

	위험도 평가 지수 개발	
논문	공정기반의 건설현장 안전 위험도 평가지수 및 위험예측 시스템 개발(박환표, 한재구, 한국건설기술연구원)	건축시공현장관리를 위한 가설공사 위험도 지수 모델 제안 (조예림, 김연철, 신윤석)
데이터	건설재해사고 통계(한국산업안전보건공단)	고용노동백서(고용노동부), 기상연보 및 기상월보
변수	안전사고 발생 빈도: 발생건수, 발생비율 안전사고 발생 강도:사망자수,부상자수, 사망자·부상자의 발생비율 위험요인: 공종, 공사종류, 공정율, 공사규모(계약금액), 층수, 공사 기간, 안전교육, 현장근무일수, 날씨(온도), 요일	기상환경요소(월별 평균기온) 작업조건요소(작업규모, 연관 작업공종, 작업시간대) 작업자조건요소(숙련도, 평균연령)
위험도 평가지수	공종별 안전사고 발생비율 + 공종별 안전사고 발생강도 비율	0.1x기상환경요소 + 0.2x작업조건요소 + 0.7x작업자조건요소
결론	공종별 위험도 평가지수 예측값 = 과거 공종별 위험도 평가지수 + 프로젝트 위험성 (날씨+공정율+공사규모+공사기간+요일)의 가중치	위험도 지수: 0~60 (안전), 60~80 (주의), 80~100 (위험)
한계점	실제 사고 데이터를 활용한 검증 부족	건설공사 중 가설공사에만 한정 재해발생형태를 요소별로 나눈 기준 불분명

# 연구방법

## 2-3 변수 확인

구분	변수명	내용	출처
종속변수	인적사고	산업재해 발생 유형(넘어짐, 떨어짐, 물체에 맞음, 등)	건설공사 안전관리 종합정보망
공사 특성	보호조치여부	조치 여부(개인방호, 안전방호)	
	공정률	10퍼센트 단위 구분	
	작업자수	작업자수 규모	
	공사기간	공사기간 종료일 - 공사시작일(전체, 공중)	
	공종	진행한 공사의 종류	
	사고객체	산재가 발생한 물적 객체	
	작업프로세스	산재가 발생한 작업 종류	
	장소	신고된 건축물의 용도	
	부위	산재가 발생한 작업 부위	
	공사비	공사비 규모(전체, 공중)	
공간 특성	시	시, 도	직접 추가
시간 특성	월	월별 구분	
	요일	요일별 구분	
날씨 특성	온도	노동부 고시 고온의 노출 기준에 따라 재분류	기상청
	풍속	풍속	
	풍향	각도로 표기	
	습도	습도	
	일조	일조량	
	전운량	구름이 덮고 있는 하늘의 비율	
	중하층운량	대기 중층,하층의 구름의 비율	
	시정	시야가 보이는 정도	

연구방법

2-4 데이터 분석 – 기초통계량

숫자형 변수

	count	mean	std	min	25%	50%	75%	max
풍속	23526	2.251152	1.449625	0	1.2	2	3	13.1
풍향	23526	185.51305	108.39393	0	90	200	270	360
습도	23526	60.440066	20.732705	0	45	60	76	100
일조	23526	0.578458	0.436851	0	0	0.8	1	1
전운량	23526	4.850973	3.854327	0	0	6	8	10
중하층운량	23526	2.779733	3.374295	0	0	1	6	10
시정	23526	2188.0838	1302.6688	0	1410	2000	2384.75	6540
총공사기간	23526	845.73327	603.39087	0	419	791	1087	9830
공종공사기간	23526	427.99277	419.13558	0	91	334	641	9830

명목형 변수

변수명	개수
요일	7
인적사고	16
보호(방호)조치여부 - 안전방호	3
보호(방호)조치여부 - 개인방호	3
공종_중분류	39
사고객체_대분류	9
사고객체_중분류	54
작업프로세스	34
장소_대분류	37
장소_중분류	6
부위_중분류	10
사고원인_주원인	29
시	18
전체공사비	18
해당공종_공사비	19
작업자수	6
공정률	10
지면온도	5
월	12

연구방법

2-4 데이터 분석 – 기초통계량

범주형 변수

보호(방호)조치여부 - 안전방호	count
조치/해당없음	17153
결측치	5215
비조치	1158

보호(방호)조치여부 - 개인방호	count
조치/해당없음	17455
결측치	5214
비조치	857

공정률	count
90% 이상	3481
40% 미만	2586
30% 미만	2586
20% 미만	2398
50% 미만	2232
60% 미만	2201
90% 미만	2155
70% 미만	2013
80% 미만	1988
10% 미만	1886

시	count
경기도	6312
서울특별시	2385
경상남도	2004
부산광역시	1365
인천광역시	1363
충청남도	1350
경상북도	1215
충청북도	1139
전라남도	1113
대구광역시	1070
강원도	913
전라북도	865
대전광역시	605
광주광역시	596
울산광역시	482
세종특별자치시	475
제주특별자치도	197
강원특별자치도	77

작업자수	count
19인 이하	7921
299인 이하	4808
49인 이하	4279
99인 이하	3123
500인 이상	1966
499인 이하	1429

요일	count
월요일	4038
화요일	4031
수요일	3924
목요일	3816
금요일	3813
토요일	2994
일요일	910

월	count
10	2380
7	2264
11	2154
8	2055
9	2015
4	1969
3	1968
12	1925
5	1893
6	1809
2	1556
1	1538

지면온도	count
25도 미만	13690
30도 초과	6107
25.9도 미만	1561
27.9도 미만	1460
30도 미만	708

사고원인_주원인	count
작업자 부주의	13030
기타	2833
작업자의 불안정한 행동	1887
작업자 통제 미흡	749
위험정보 미제공	650
작업순서 미준수	595
운전자 자격관리 미흡	417
시공품질 미확보	411
시공관리 미흡	371
시방기준 등 미준수	299
개인 안전보호구 착용 불량	292
부적정 기계장비 사용 통제 미흡	196
사전 설계도서검토 미흡	181
안전관리계획 수립 미흡	172
구조안전 미확보	154
불량자재 관리 미흡	147
작업자 건강상태 확인 미흡	139
작업전 점검 미흡	138
기계장비 불안정한 거치	120
개인 안전보호구 미착용	105
안전발판 설치 미흡	89
적정 공법검토 미흡	89
계측데이터 관리 미흡	81
설치·해체과정 관리 미흡	77
단독작업	71
사다리 사용수칙 미준수	67
품질관리 미흡	65
지반조사 미흡	53
구조계산 오류	48

연구방법

2-4 데이터 분석 – 기초통계량

범주형 변수

해당공종_공사비	count	전체공사비	count	공종_중분류	count	작업프로세스	count	작업프로세스	count
20억~50억원미만	3025	1,000억원 이상	7944	철근콘크리트공사	7066	설치작업	4125	정비작업	112
10억~20억원미만	2375	500억 ~ 1,000억원 미만	2650	가설공사	2542	해체작업	2525	반출작업	101
50억~100억원미만	2110	50억 ~ 100억원 미만	1994	기타	2077	이동	2521	별목작업	96
5억~10억원미만	1904	20억 ~ 50억원 미만	1873	해체 및 철거공사	1044	운반작업	2071	항타 및 항발작업	93
분류불능	1901	300억 ~ 500억원 미만	1788	기계설비공사	952	기타	1846	거치작업	81
100억~150억원미만	1576	100억 ~ 150억원 미만	1526	토공사	917	정리작업	1550	적재작업	76
1,000만원미만	1519	200억 ~ 300억원 미만	1407	철골공사	778	조립작업	1131	옹벽	94
1억~2억원미만	1212	10억 ~ 20억원 미만	1010	건축물 부대공사	643	형틀 및 목공	933	천공기	89
4,000만~1억원미만	1091	150억 ~ 200억원 미만	977	수장공사	598	타설작업	794	작업대차	89
200억~300억원미만	1086	5억 ~ 10억원 미만	735	관공사	564	마감작업	785	지지대	85
3억~5억원미만	1024	3억 ~ 5억원 미만	460	건축 토공사	553	절단작업	687	특수거푸집(갱폼 등)	75
150억~200억원미만	942	1억 ~ 2억원 미만	265	미장공사	536	준비작업	595	핀	72
300억~500억원미만	873	2억 ~ 3억원 미만	251	전기설비공사	518	설비작업	364	벽체	72
2억~3억원미만	741	4,000만 ~ 1억원 미만	192	타일 및 돌공사	458	청소작업	303	전주, 전선	71
1,000억원이상	585	분류불능	115	도로 및 포장공사	397	도장작업	286	콘크리트믹서트럭	71
500억~1,000억원미만	531	1,000만 ~ 2,000만원 미만	115	교량공사	344	굴착작업	273	띠장	68
2,000만~4,000만원미만	525	2,000만 ~ 4,000만원 미만	114	조적공사	333	연결작업	250	지하매설물	67
1,000만~2,000만원미만	506	1,000만원 미만	110	도장공사	301	상차 및 하역작업	247	절토사면	65
				터널공사	290	용접작업	246	비산물	64
				산업설비공사	287	인양작업	235	경사면	64
				목공사	277	양중작업	169	버팀보	57
				금속공사	270	확인 및 점검작업	168	천정패널	57
				창호 및 유리공사	245	쌓기작업	160	개구부	55
				하천공사	227	매설작업	156	덕트	55
				방수공사	210	보수 및 교체작업	155	버팀대	52
				관공사 부대공사	198	천공작업	149	와이어로프	51
				항만공사	161	전기작업	130		
				지붕 및 홀통공사	156	부설 및 다짐작업	113		
				철도 및 궤도공사	122				
				말뚝공사	117				
				지정공사	93				
				조경공사	64				
				지반개량공사	49				
				강구조물공사	45				
				지반조사	35				
				프리캐스트 콘크리트공사	21				
				댐 및 제방공사	20				
				특수 건축물공사	9				
				통신설비공사	9				

연구방법

2-4 데이터 분석 – 기초통계량

범주형 변수

사고객체_대분류	count
가시설	6643
기타	5152
건설자재	4568
건설공구	2468
건설기계	1846
부재	978
시설물	968
토사 및 암반	540
질병	363

사고객체_중분류	count
기타	5477
자재	2783
공구류	2008
거푸집	1844
비계	1366
철근	1024
건물	799
작업발판	758
기타 가시설	552
흙막이가시설	449
굴착기	446
사다리	445
시스템동바리	432
질병	363
철골부재	319
강관동바리	251
안전시설물	223
기중기(이동식크레인 등)	216
슬래브	202
배관	200
고소작업차(고소작업대 등)	195
차량	182
가설계단	181
덤프트럭	161
데크플레이트	160
지반	149
타워크레인	139

사고객체_중분류	count
항타 및 항발기	136
파이프서포트	133
굴착사면	133
지게차	113
창호	111
건설폐기물	106
콘크리트펌프	97
옹벽	94
천공기	89
작업대차	89
지지대	85
특수거푸집(깁폼 등)	75
핀	72
벽체	72
전주, 전선	71
콘크리트믹서트럭	71
띠장	68
지하매설물	67
절토사면	65
비산물	64
경사면	64
버팀보	57
천정패널	57
개구부	55
덕트	55
버팀대	52
와이어로프	51

장소_대분류	count
공동주택	7027
기타	2481
공장	1920
업무시설	1776
교육연구시설	1492
근린생활시설	1424
도로	917
창고시설	723
하수도	563
부지조성	484
문화및집회시설	453
상수도	382
숙박시설	347
도로교량	340
단독주택	300
운동시설	243
의료시설	231
교정및군사시설	228
자동차관련시설	196
노유자시설	186
판매시설	178
석유화학공장	163
지하철	156
도로터널	155
종교시설	148
철도터널	144
운수시설	143
제방(통관/호안)	109
옹벽	105
일반및고속철도	89
하수처리시설	88
방파제	64
방송통신시설	63
절토사면	61
관개수로	50
위험물저장및처리시설	49
계류시설	48

장소_중분류	count
내부	10552
외부	6467
기타	4874
인접주변	999
외벽	365
지붕	269

부위_중분류	count
바닥	7371
기타	6542
상부(위)	3801
하부(아래)	1333
옆	1302
지하	1255
고소	937
앞	693
뒤	219
계단	73

## 연구방법

### 2-5 데이터 전처리

#### 범주형 데이터 전처리

범주형 변수 중 50종 이상의 고유값을 갖는 변수 중 비율이 1%미만인 항목은 '기타/분류불능'으로 재분류

작업프로세스	
설치작업	4148
이동	2540
해체작업	2539
운반작업	2091
정리작업	1554
기타	1526
조립작업	1137
형틀 및 목공	936
타설작업	805
마감작업	787

작업프로세스	
정지작업	2
점검	2
제거작업	2
돌붙임 및 줄눈 작업	2
수정작업	2
다짐작업	2
평탄작업	1
교체작업	1
토사하역	1
전기배선작업	1

#### 숫자형 데이터 전처리

총_공사기간	
count	23722.00000
mean	845.38791
std	604.05017
min	-2062.00000
25%	419.00000
50%	791.00000
75%	1089.00000
max	9830.00000

총_공사기간	
count	23526.00000
mean	845.733274
std	603.390871
min	0.000000
25%	419.000000
50%	791.000000
75%	1087.000000
max	9830.000000

기온	
count	23526.000000
mean	20.900897
std	13.229914
min	-12.700000
25%	10.400000
50%	22.300000
75%	30.200000
max	63.200000

기온	
count	23526.000000
mean	21.085174
std	31.566378
min	-83.000000
25%	10.000000
50%	18.000000
75%	25.000000
max	436.000000



## 2-5 데이터 전처리 - 차원 축소

## PCA

기상데이터에 대한 연속형 변수는 PCA를 통하여 1개의 차원으로 축소 가능, 1개 Column으로 기존 컬럼의 설명된 변수를 0.99만큼 설명할 수 있음

PC1
-139.767347
-242.810174
-1081.846957
-139.767347
-136.782131
...



습도	기온.1	강수량	풍속	습도.1
0.60	16.8	0.0	1.0	58.0
0.55	17.0	0.0	1.0	66.0
0.80	16.6	0.0	1.5	76.0
0.60	16.8	0.0	1.0	58.0
0.52	21.7	0.0	2.4	61.0
...	...	...	...	...

```
[pca] >Number of components is [1] that covers the [99.00%] explained variance.
```

## Feature Selection

1차적으로 학습한 모델의 Feature Importance를 출력하여 해당 값이 0.001이상인 Feature만을 2차 학습에 사용하여 차원 축소를 적용, 모델 학습의 효율과 성능을 개선할 수 있음

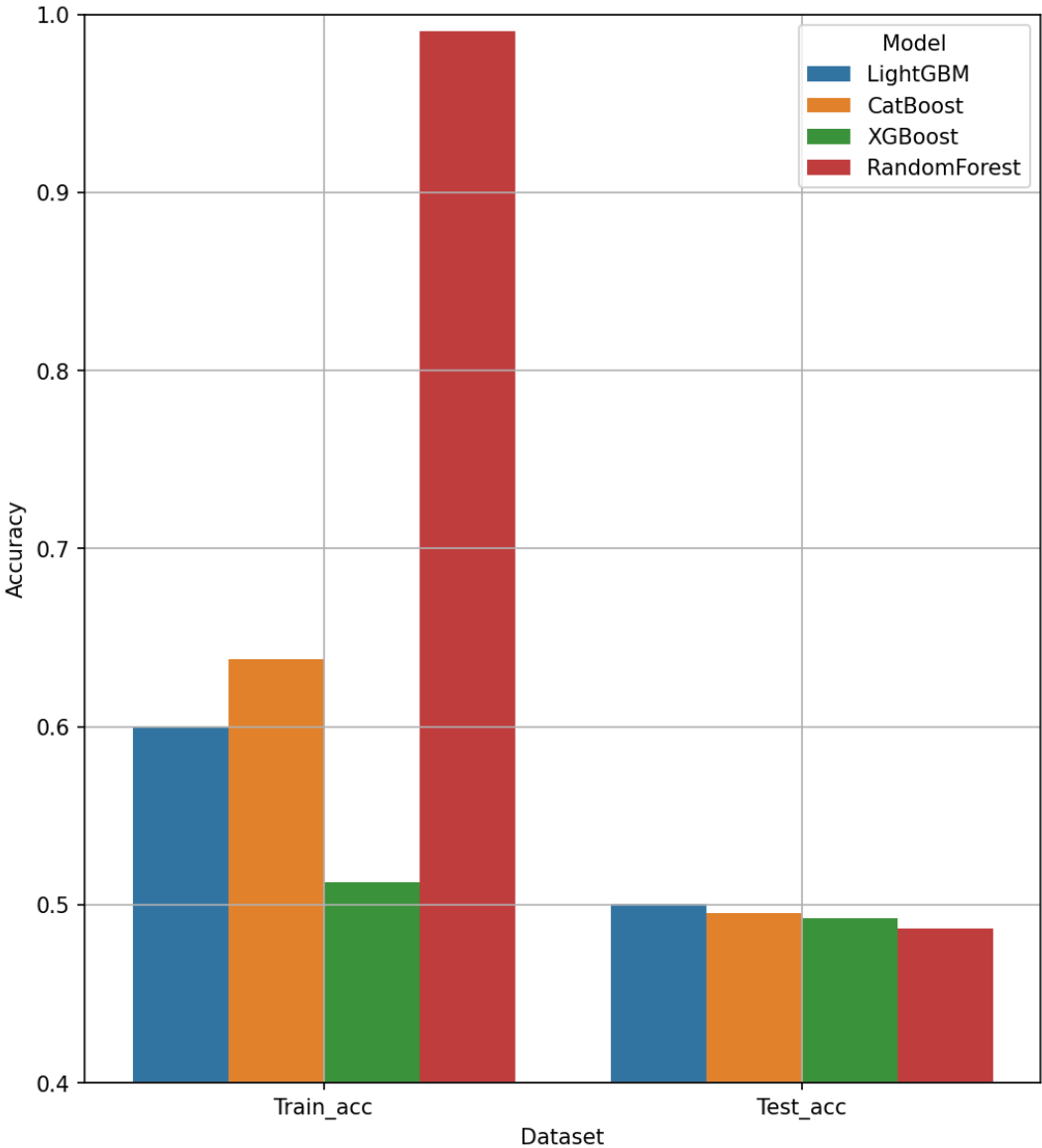
Column	Feature_Importance
보호(방호)조치여부 - 개인방호	0.039558
사고객체_중분류_공구류	0.034327
부위_중분류_고소	0.026299
부위_대분류_공구류	0.025992
작업프로세스_이동	0.022276
...	...

단일 모델 성능 측정

1. RandomForest, XGBoost, LightGBM, CatBoost 총 4종의 모델을 대상으로 학습 및 성능 평가 진행
2. 각 모델은 GridsearchCV를 사용한 하이퍼파라미터 튜닝을 진행, cv = 10으로 각 모델 성능 교차검증을 진행함
  - 모델의 하이퍼파라미터 튜닝 : GridsearchCV를 활용한 하이퍼파라미터별 성능 탐색
  - Cross validation으로 10분할 k-fold를 수행하여 과적합 방지
  - 성능 평가 지표는 정확도(Accuracy)를 사용
  - Validation set의 Accuracy를 평가 기준으로 선택하되 Train set의 Accuracy를 과적합 판별에 참고함
3. 개별 모델 훈련 결과 LightGBM에서 가장 우수한 성능을 확인함  
RandomForest 모델의 경우 훈련 데이터에 대한 과대적합이 발생

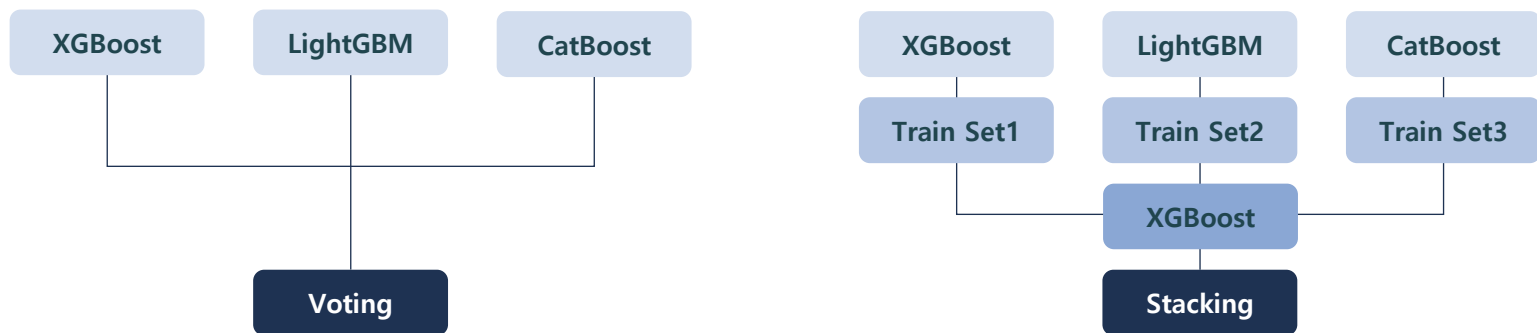
Model	Train_acc	Test_acc
LightGBM	0.599681	0.500425
CatBoost	0.637938	0.495113
XGBoost	0.512593	0.492138
RandomForest	0.990755	0.486400

Column	Feature_Importance
보호(방호)조치여부 - 안전방호	7.855503
보호(방호)조치여부 - 개인방호	7.678744
공종공사기간	5.278317
총_공사기간	5.019284
시정	3.954574



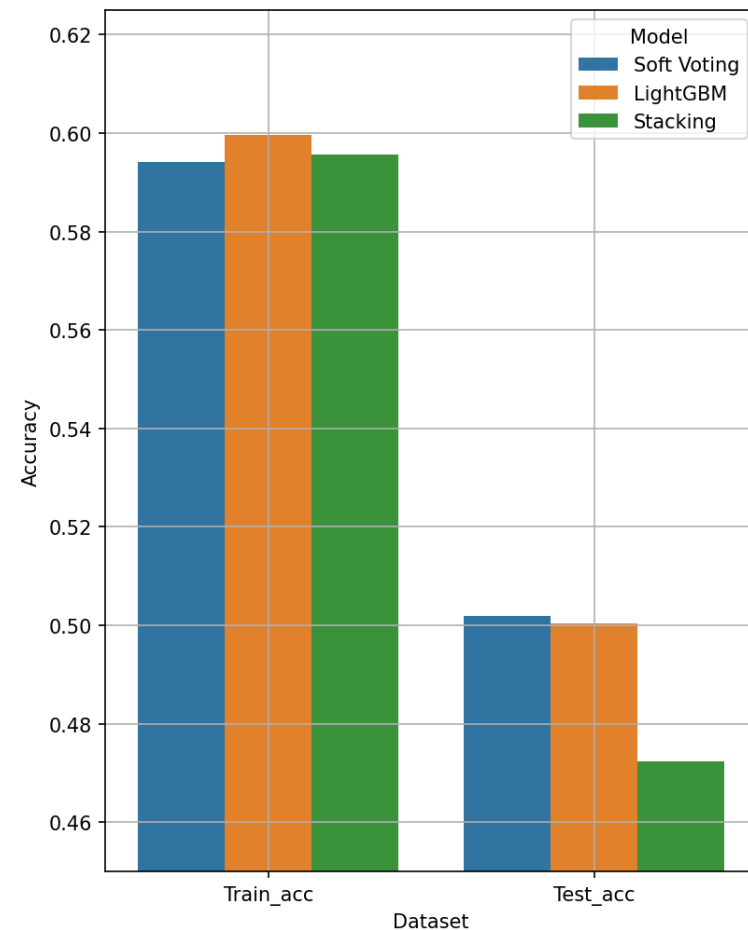
### 앙상블 모델 성능 측정

1. 목적 : 단일 모델로서 일정 이상의 성능을 나타내는 모델들을 앙상블하여 추가적인 성능 향상이 가능한지의 여부를 확인
2. 각 모델의 성능을 확인한 결과 XGBoost, LightGBM, CatBoost 3종의 모델을 사용하기로 결정
3. 다수의 다른 종류의 모델을 앙상블하는 방법으로 Voting과 Stacking을 선정
  - Stacking의 Final estimator는 XGBoost를 사용



4. 앙상블 결과 (Soft) Voting에서 모델 성능이 LightGBM 단일 모델 성능보다 우수한 결과를 보임

	Train_acc	Test_acc	Type
Model			
Soft Voting	0.594155	0.501912	Ensemble
LightGBM	0.599681	0.500425	Single
Stacking	0.595696	0.472376	Ensemble



## 종속변수 분리 모델 성능 측정

1. 목적 : 변수를 특정 기준에 따라 분리하여 모델링을 별도로 수행하여 모델 성능 향상여부를 확인
2. 변수를 분리한 모델의 정확도가 개선될 수 있으나 임의 변수 예측 시 어떤 모델을 선택할 것인지에 대한 정확도를 추가로 고려할 필요가 있음

- 모델 A 종속변수

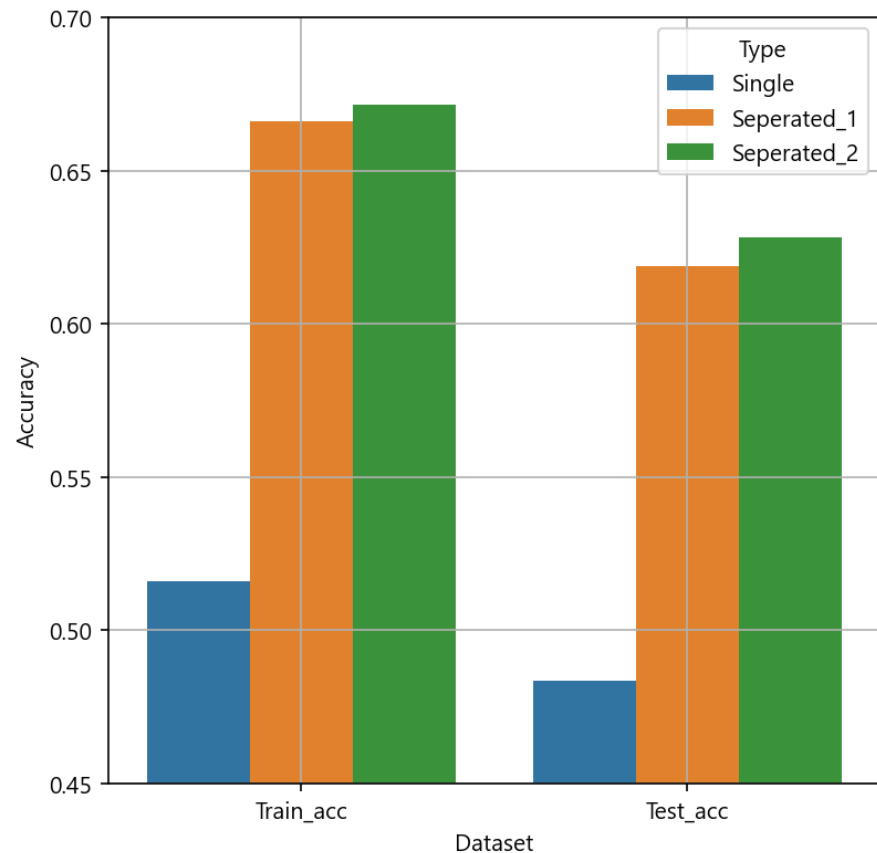
['넘어짐', '떨어짐', '부딪힘', '절단, 베임', '칼림']

- 모델 B 종속변수

['물체에 맞음', '끼임', '질병', '찔림', '화상', '교통사고', '감전', '기타', '질식']

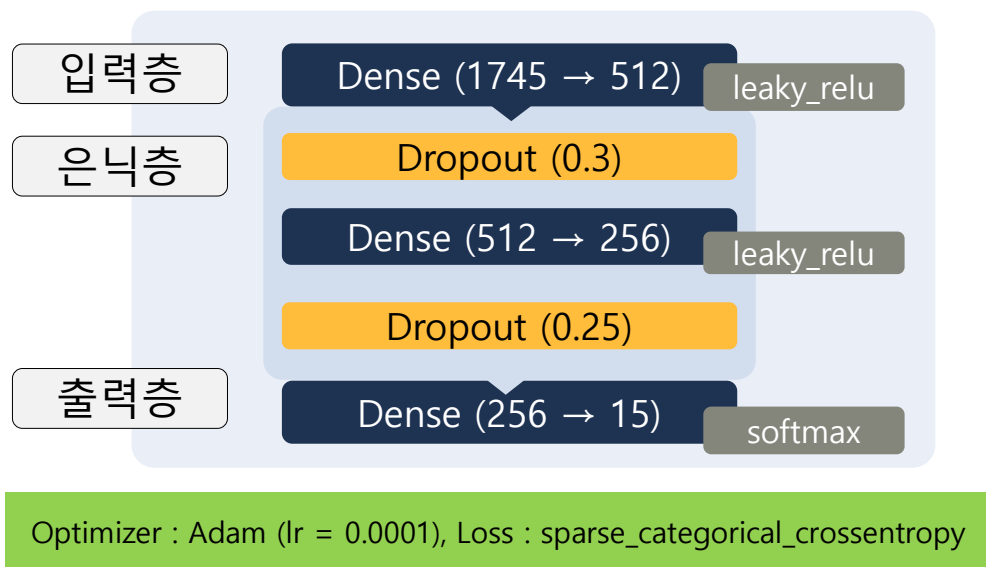
3. 각 모델의 정확도는 변수 미분리 모델의 정확도보다 높음

Train_acc	Test_acc	Type
0.515995	0.483743	Single
0.666138	0.618880	Seperated
0.671657	0.628134	Seperated

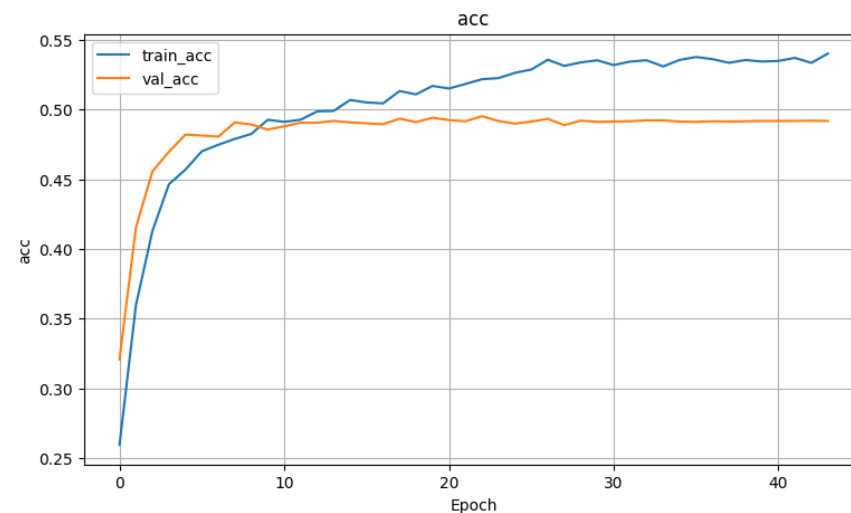
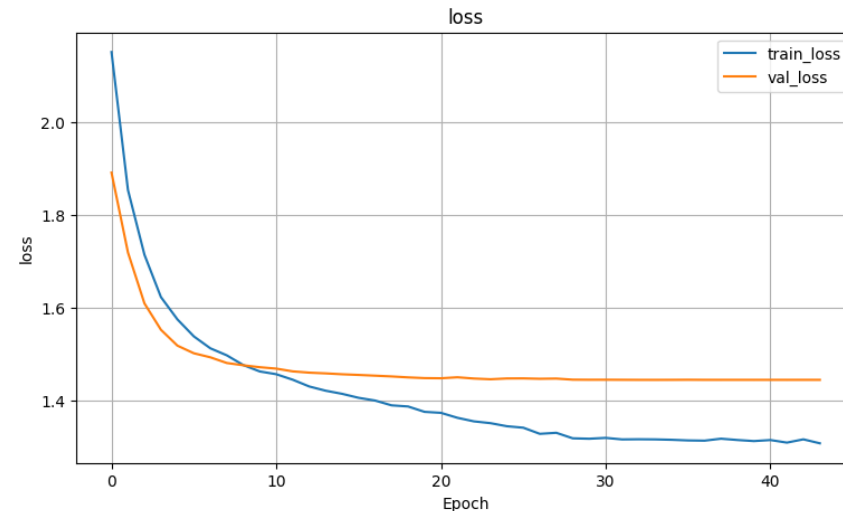


### 딥러닝 모델 구현 및 성능 평가

1. 딥러닝 모델을 구현하기 위한 분석 목표 확인
  - 여러 컬럼의 데이터로 구성된 다중 분류 문제
  - 은닉층의 구성은 Dense 층과 Dropout을 사용하여 적층
  - 출력층은 softmax 활성화 함수를 적용한 단일 Dense 층으로 구성
2. 딥러닝 모델 훈련을 위한 모델 구성 및 하이퍼파라미터 선정



3. 딥러닝 모델 학습 결과 Train acc 0.56, Validation acc 0.495로 확인



## 연구 결과

1. 모델의 변수중요도를 확인한 결과 개인 보호 조치 여부와 작업 프로세스가 건설업에서 산업재해 발생 유형을 결정하는데 중요한 요소임을 알 수 있음
2. 산업재해 발생 유형을 정확도 0.5 내외로 예측하는 모델 완성
  - XGBoost, LightGBM, CatBoost 3종의 모델을 Soft Voting 앙상블을 수행한 모델의 성능이 가장 우수함
3. 딥러닝 모델의 정확도가 0.495로, 본 연구에서 제시한 모델이 낮은 정확도를 갖는 모델이 아님을 알 수 있음

## 연구의 시사점

1. 특정한 조건(공사 현장)하에서 발생가능성이 높은 산재 유형을 추정하여 산업재해 예방에 기여할 수 있음
2. 산업재해를 발생시키는 요인은 많은 요소에 의하여 복합적으로 작용하며, 그에 따라 발생할 수 있는 산업재해도 한 가지에 국한되지 않음
3. 산재 발생 유형과 관련 없이 공통적으로 개인 방호 조치에 대한 중요도는 높아 이에 대한 대비가 중요함을 시사함

## 결론

# 4-2 정책 제언 및 연구 기대 효과

## 정책 제언

1. 발생 유형에 따른 각각의 예방대책방안 수립
2. 해당공사현장에서 발생 가능성이 높은 3가지 유형을 제공하고 경각심을 제고
3. 개인별 부주의가 산업 재해 발생원인중 대다수(64.8% 이상)이므로  
작업자 통제 강화

## 연구 기대효과

1. 건설업 환경에서의 고위험 요소 사전 감지 및 사전 예방 가능
2. 정부 / 기업 / 구직자에 대한 기대 효과 제언...
  - 정부는 고위험군 발생 예방을 위한 법적 정책 수립
  - 기업은 산재 사전 발생을 대비하여 이윤 손실 최소화
  - 구직 / 종사자는 직업 선택 및 산재 발생에 의한 책임 부가에 대한 전문적인 뒷받침이 되는 자료 제공

### 일일 재해위험 안내

#### 입력 데이터

지역 서울특별시 강남구

요일 금요일(평일)

현재 시각 날씨



온도: 20 °C

습도: 60 %

풍속: 3 m/s

공사명

○○건설 XX빌딩 재건축

공사기간: 420일

공사비: 300억 원

작업자수: 400명

공정률: 30 %

보호조치여부: 조치

공종: 가설공사

#### 주의해야 할 재해유형



떨어짐 40%



넘어짐 32%



물체에 맞음 25%



## 4-3 연구의 한계점 및 향후 연구제안

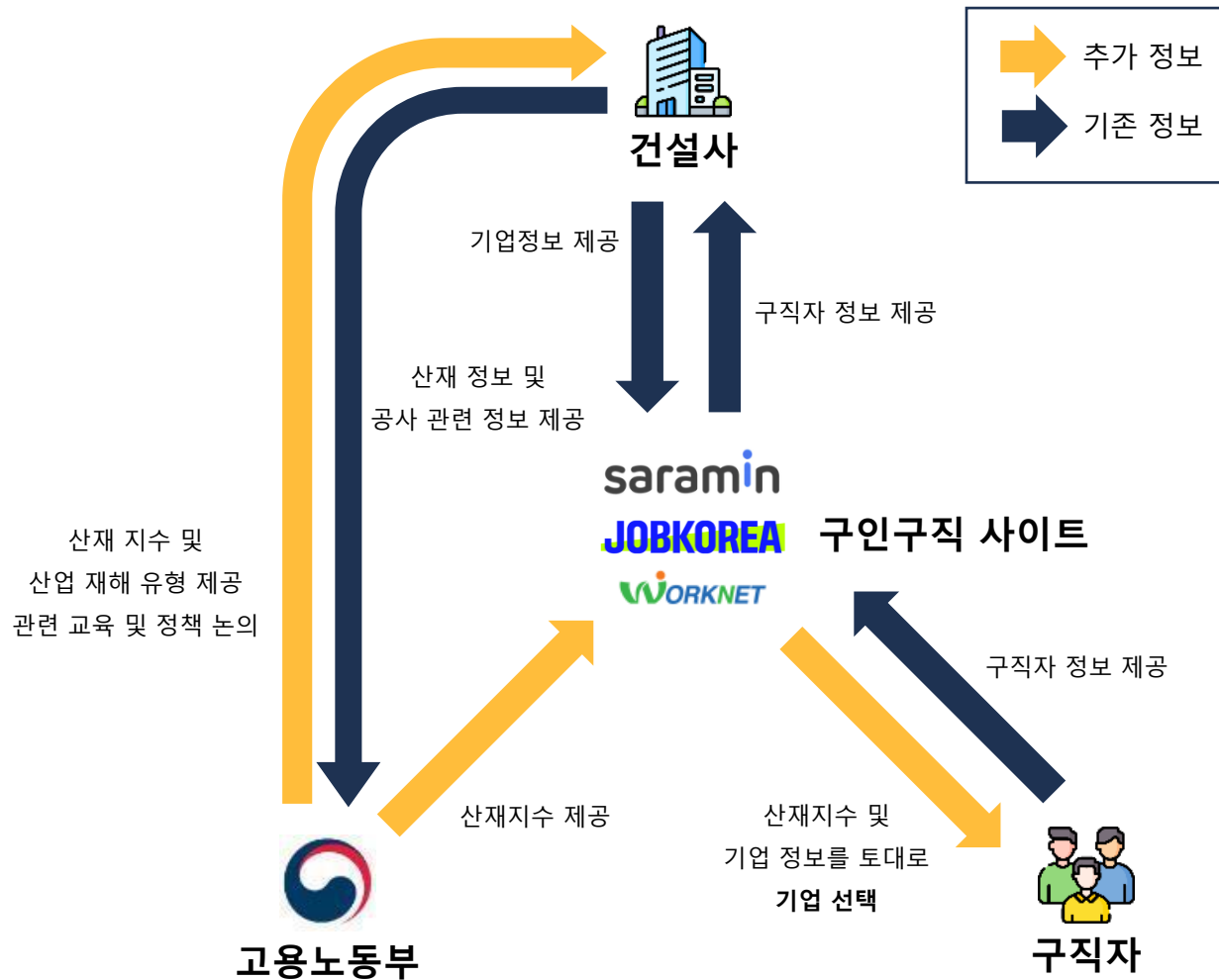
### 연구의 한계점

1. 산재 발생을 자체는 확인하기 어려움
2. 0.5의 정확도에서 추가 개선 필요
3. 개인에 대한 특성이 데이터에 존재하지 않음
4. 수집한 자료 중 텍스트에 해당하는 부분의 텍스트 마이닝 결과를 제시하지 않음

### 향후 연구 제안

향후 고도화한 예측모델로 산재 지수를 제공하여 정부부처와 구인 기업, 구직자에게 이익을 극대화하고자 함

- **구직자** : 지원을 희망하는 기업의 안전에 대한 관심도와 그에 따른 기업 내 정책을 살펴봄에 기업 선택의 기준으로 삼음
- **기업** : 산업재해 예방에 따른 기업의 이미지 관리로 브랜드 이미지 제고 및 기업이 희망하는 구직자 확보 가능
- **정부** : 기업의 산업재해에 대한 자체적인 관리를 유도하여 사회적인 비용 감소



- e-나라지표. n.d. “산업재해현황”, 2024년 5월 9일 갱신, [https://www.index.go.kr/unity/potal/main/EachDtlPageDetail.do?idx\\_cd=1514](https://www.index.go.kr/unity/potal/main/EachDtlPageDetail.do?idx_cd=1514)
- KOSIS. n.d. “전체 재해 현황 및 분석-업종별(산업별 중분류)”, 2024년 1월 18일 갱신, [https://kosis.kr/statHtml/statHtml.do?orgId=118&tblId=DT\\_11806\\_N022&vw\\_cd=MT\\_ZTITLE&list\\_id=118\\_11806\\_ciek6458&seqNo=&lang\\_mode=ko&language=kor&obj\\_var\\_id=&itm\\_id=&conn\\_path=MT\\_ZTITLE](https://kosis.kr/statHtml/statHtml.do?orgId=118&tblId=DT_11806_N022&vw_cd=MT_ZTITLE&list_id=118_11806_ciek6458&seqNo=&lang_mode=ko&language=kor&obj_var_id=&itm_id=&conn_path=MT_ZTITLE)
- 안전보건공단. n.d. “「산업재해통계 인포그래픽」- 통계로 보는 2022년도 산업재해”, 2023년 3월 29일, <https://www.kosha.or.kr/kosha/data/mediaBankMain.do>
- 김학열, 허태영, “산업재해 발생의 상대위험도 분석 및 순환분포 모형 추정,”(서울도시연구 제11권 제1호, 2010. 3), 127~138.
- 김현수, 이현수, 박문서, “건설공사의 정량적 위험도 산정 방법론,”(한국건설관리학회 학술발표대회 논문집, 2008), 463~466
- 조민건, 이동환, 박주영, 박승희, “건설현장 정형·비정형데이터를 활용한 기계학습 기반의 건설재해 예측 모델 개발,”(대한토목학회 논문집 제42권 제1호(통권 제220호), 2022), 127~134
- 김명중, 박선영, “기업 특성이 산업재해 발생에 미치는 영향 : 중소기업과 대기업 비교,”산업연구 Journal of Industrial Studies(J.I.S)(2023) Vol.4 No.2
- 서동민, “인공지능을 활용한 산업재해 예방 현황과 전망,” 한국콘텐츠학회(2023) Vol.21 No.1
- 송태호, “건축시공현장관리를 위한 가설공사 위험도 지수 모델 제안,” (석사학위, 금오공과대학교 산업대학원 토목,환경 및 건축공학과, 2019)
- 박환표, 한재구, “공정기반의 건설현장 안전 위험도 평가지수 및 위험예측 시스템 개발(Ⅰ),”(한국건설기술연구원 2018년도 주요사업 연차보고서(창의·시드사업),2018),KICT 2018-076
- 조예림, 김연철, 신윤석, “의사결정나무기법을 이용한 건설재해 사전 예측모델 개발,”(한국건축시공학회지 제17권 제3호, 2017), 295~303