**건설안전사고 자동분류 시스템**

**학번 : 2218065**

**이름 : 박성아**

**Github address :**

1. **안전 관련 머신러닝 모델 개발의 목적** 
   1. 학습 모델 활용 대상 : 어떠한 독립 변수를 사용하여 안전사고 여부를 예측하고자 함.
   2. 개발의 의의 :

-안전사고를 자동으로 분류하는 모델을 개발하여 현장에서의 안전성을 향상시킬 수 있음.

- 사고 발생을 사전에 감지하여 조치를 취함으로써 작업자의 안전을 보장하고 재무적 손실을 최소화할 수 있음.

- 데이터 기반의 예측 모델을 활용함으로써 미래의 사고 발생 가능성을 예측하고 예방할 수 있음.

- 안전 관리 및 감독 업무를 효율적으로 수행할 수 있는 지표 제공으로 안전 관리체계를 강화할 수 있음.

1. **안전 관련 머신러닝 모델의 네이밍의 의미** 
   1. 예) 개똥이 모델은 ~~ 의미가 있다.
2. **개발 계획**
   1. 데이터에 대한 요약 정리 및 시각화

: - 데이터 특성 및 구조를 파악하기 위해 요약 통계 및 시각화를 수행.

- 클래스 불균형 여부를 확인하고, 안전사고와 비안전사고 클래스의 분포를 시각적으로 파악.

* 1. 데이터 전처리 계획

: - 이미지 데이터인 경우: 크기 조정, 정규화, 데이터 증강 등을 수행하여 모델 학습에 적합한 형태로 가공.

- 불필요한 변수나 결측치 처리 등 다양한 전처리 기법을 사용.

* 1. 어떠한 머신러닝 모델을 사용할 것인지 (해당 머신러닝 모델의 이론 추가)

: - CNN (Convolutional Neural Network) 모델 사용.

- CNN은 이미지 분류에 효과적이며, 공간적인 특징을 잘 학습할 수 있는 모델.

이론 :

- CNN은 여러 개의 합성곱(Convolution)과 풀링(Pooling) 레이어로 구성됨.

- 이미지의 지역적인 특징을 학습하고, 이를 기반으로 이미지를 분류하는 데 사용.

* 1. 머신러닝 모델 예측 결과:

- 모델이 안전사고 또는 비안전사고로 분류할 것으로 예상.

- 모델의 예측 확률을 통해 사고 발생 가능성을 확인할 수 있음.

* 1. 사용할 성능 지표

: - 정확도 (Accuracy): 전체 샘플 중 모델이 정확하게 예측한 비율.

- 정밀도 (Precision): 모델이 안전사고로 예측한 것 중 실제 안전사고인 비율.

- 재현율 (Recall): 실제 안전사고 중 모델이 정확하게 예측한 비율.

- F1 점수: 정밀도와 재현율의 조화 평균.

* 1. 성능 검증 방법 계획

- 교차 검증 또는 홀드아웃 검증을 사용하여 모델의 일반화 성능을 평가.

- 혼동 행렬(Confusion Matrix)을 통해 각 클래스에 대한 모델의 성능을

시각화하고 평가.

1. **개발 과정** 
   1. 계획 후 실제 학습 모델 개발 과정을 기록 (\*개발 과정 캡쳐 필수)
   2. 각 함수는 어떻게 동작하는 지 구체적으로 설명
   3. 에러 발생 지점 및 해결 과정
   4. 학습 모델의 성능 평가
   5. 결과 시각화
2. **개발 후기**
   1. 개발 후 느낀 점 설명