



인공지능 멘토링 [제조 분야]

이 민 태



머신러닝 (제조분야)

제조분야의 머신러닝

- 선형 회귀 분석
- 의사결정트리
- 랜덤 포레스트
- K-means

머신러닝 예제 (제조분야)

공정 개선을 위한 선형 회귀분석 - 예측

1. 핵심 질문: 무엇이 당신의 결과를 결정합니까?



"온도? 속도? 압력?"

문제의 원인을 '직감'에 의존하고 있나요?



고품질 완제품

데이터는 명확한 답을 알고 있습니다.

머신러닝 예제 (제조분야)

공정 개선을 위한 **선형 회귀분석** - 예측

2. 도구 이해하기: '조절 손잡이'와 '결과 계기판'

선형 회귀분석은 '원인'과 '결과' 사이의 관계를 명확한 직선으로 설명합니다. 우리가 직접 제어할 수 있는 '원인'을 바꾸면, 그에 따라 측정되는 '결과'가 어떻게 변하는지 예측 가능한 규칙을 찾아냅니다.



독립 변수 (X) - "원인"

(우리가 제어하는 '조절 손잡이')

- 기계 속도
- 온도, 압력
- 냉각 시간



종속 변수 (Y) - "결과"

(우리가 주시하는 '결과 계기판')

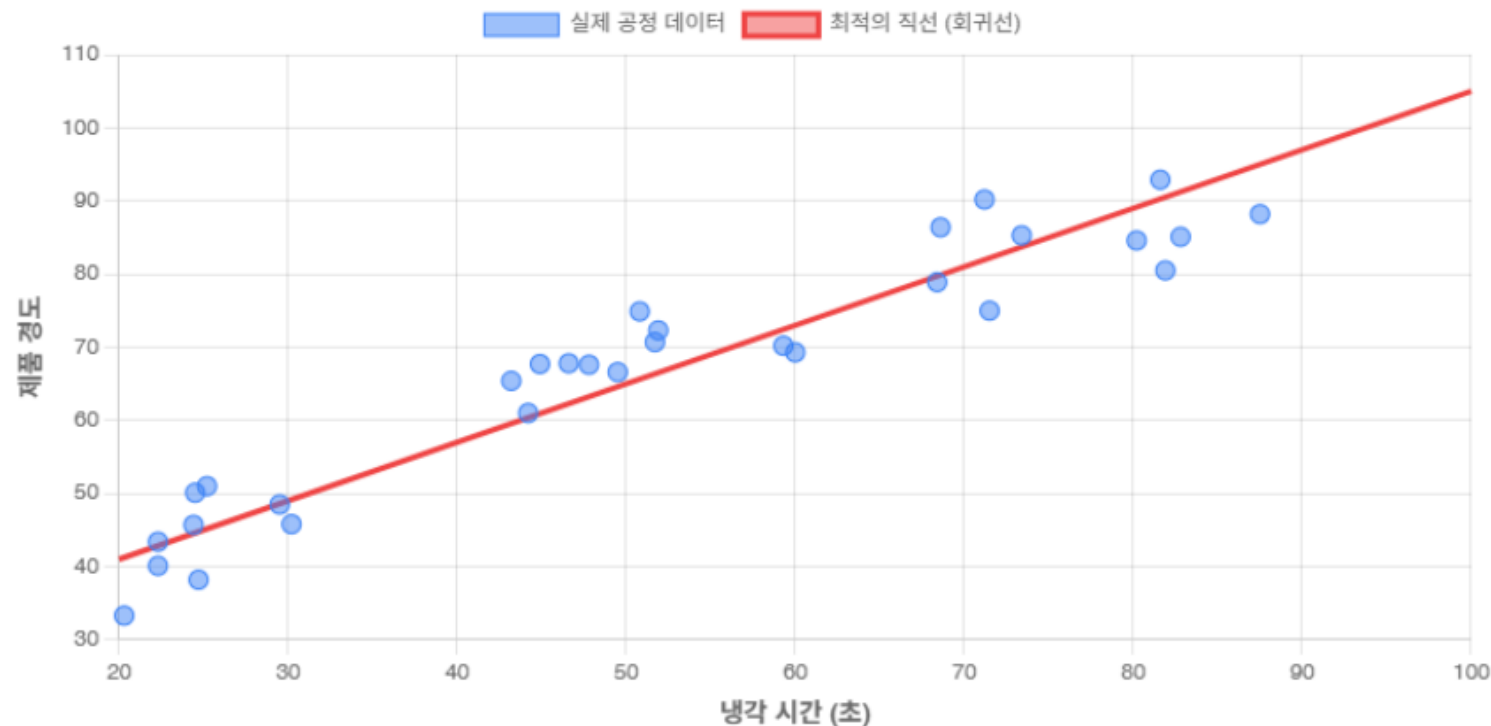
- 제품 강도
- 불량률, 수율
- 표면 조도

A 머신러닝 예제 (제조분야)

공정 개선을 위한 **선형 회귀분석** - 예측

3. 3단계 프로세스 실행: 제품 경도 최적화 예시

자동차 부품의 '경도'를 최적화하기 위해 '냉각 시간'을 조절하는 실제 사례를 살펴봅시다. 과거 데이터를 분석하여 두 변수 간의 숨겨진 규칙을 찾아냅니다.



위 차트는 냉각 시간(X축)이 증가할수록 제품 경도(Y축)가 어떻게 변하는지 보여주는 산점도와 최적의 직선입니다.

실행 가능한 규칙 (방정식)

$$\text{경도} = 0.8 * \text{시간} + 25$$

핵심 통찰 (영향력 계수)

0.8

냉각 시간 1초당 경도 0.8 증가

신뢰도 점수 (R-제곱)

0.92

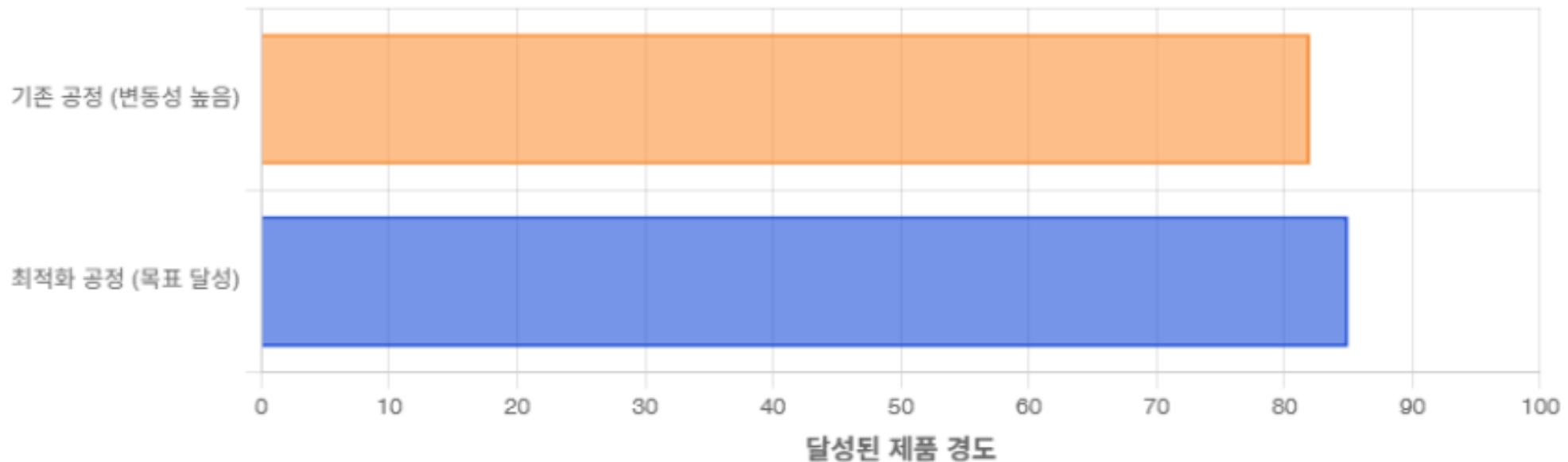
경도 변화의 92%가 설명됨

머신러닝 예제 (제조분야)

공정 개선을 위한 **선형 회귀분석** - 예측

예측: 목표 경도 85 달성하기

위 규칙을 이용하면 목표 경도 85를 달성하기 위한 최적의 냉각 시간이 75초임을 계산할 수 있습니다. 이제 추측이 아닌 데이터에 기반하여 공정을 설정할 수 있습니다.



A 머신러닝 예제 (제조분야)

공정 개선을 위한 **선형 회귀분석** - 예측

관계 속에서 '예측 공식'을 찾다

경향선 선택: 직선 vs 곡선

1. 직선 회귀: 단순함의 미학

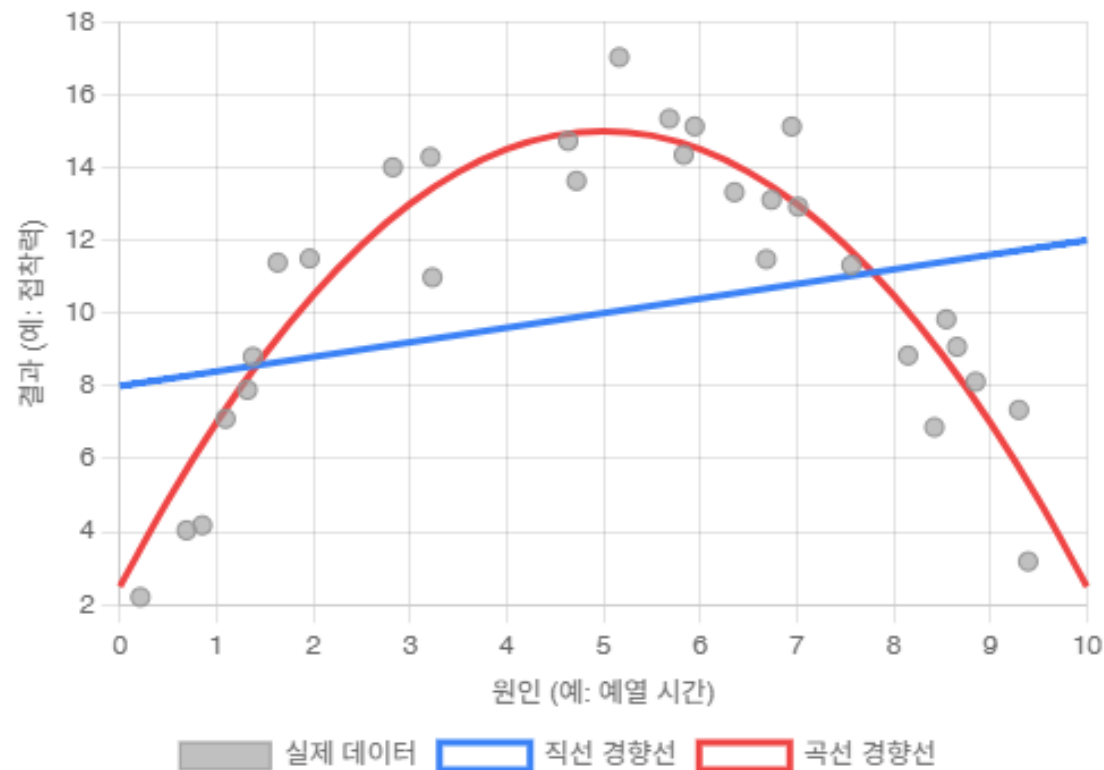
원인과 결과가 ****일정한**** 관계를 가질 때 사용. 해석이 쉽고 명확합니다.

$$\text{결과} = a \times \text{원인} + b$$

2. 곡선 회귀: 유연함의 힘

원인의 효과가 구간마다 달라지는 ****복잡한**** 패턴을 표현할 때 효과적입니다.

$$\text{결과} = a \times \text{원인}^2 + b \times \text{원인} + c$$



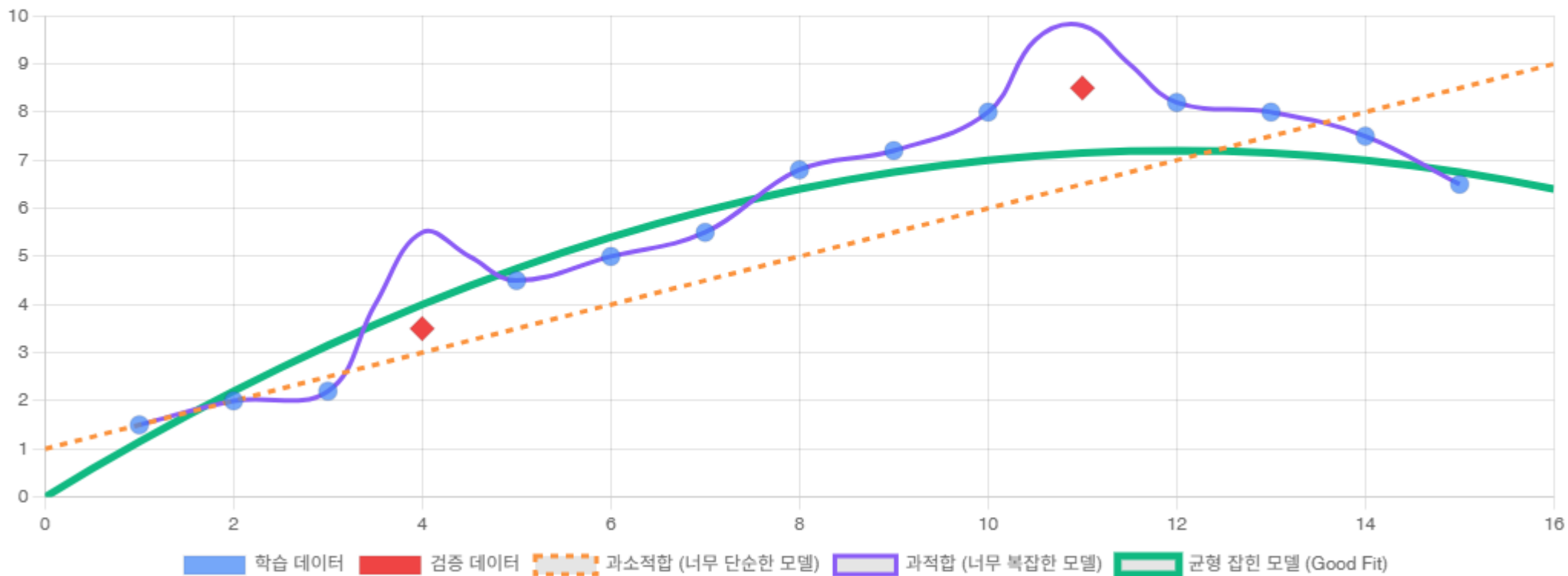


머신러닝 예제 (제조분야)

공정 개선을 위한 선형 회귀분석 - 예측

가장 큰 함정: 과적합(Overfitting)

모델이 '학습 데이터'만 너무 열심히 외워서, 처음 보는 '검증 데이터'에 대한 예측을 실패하는 현상입니다.





머신러닝 예제 (제조분야)

공정 개선을 위한 **선형 회귀분석** - 예측

4. 성과: 당신이 얻는 것



공정 최적화

품질 목표를 정밀하게 달성하고 일관성을 유지하여 고객 만족도를 높입니다.



낭비 및 비용 절감

불량품과 재작업을 획기적으로 줄여 실질적인 비용을 절감합니다.



미래 결과 예측

공정을 시작하기 전에 결과를 예측하여 더 나은 의사결정을 내립니다.



머신러닝 예제 (제조분야)

의사결정트리 - 분류

문제: 복잡하게 얽힌 원인, 어디부터 봐야 할까?



경험과 직감

"온도? 습도? 작업 시간?
무엇이 진짜 원인인지 불확실하다."



의사결정 트리

데이터가 가장 가능성 높은 원인부터 순서대로 길을 안내하는 '문제
해결 지도'

A

머신러닝 예제 (제조분야)

의사결정트리 - 분류

열처리온도 기준 (80도 이상)	정상여부
75	불량
90	정상
100	정상
73	불량
105	정상

열처리온도 기준 (80도 이상)	열처리시간 기준(30초이상)	습도 기준(60 이하)	정상여부
65	28	62	불량
80	29	58	정상
90	30	68	불량
63	29	59	정상
95	31	55	정상

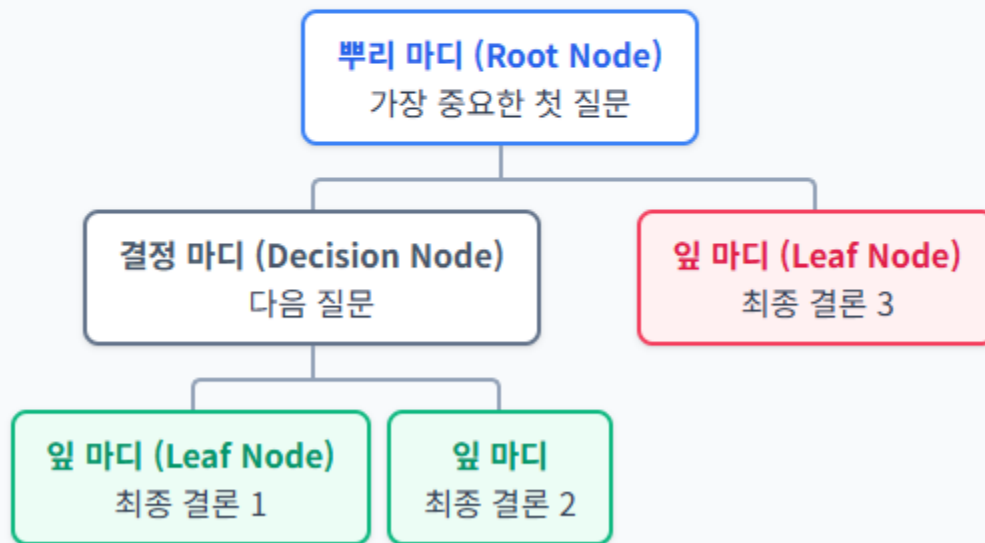


머신러닝 예제 (제조분야)

의사결정트리 - 분류

의사결정 트리란? 데이터로 하는 '스무고개 게임'

'정상'과 '불량'을 가장 잘 나눌 수 있는 결정적인 질문을 계속해서 던져, 문제의 원인을 찾아가는 과정입니다.



A 머신러닝 예제 (제조분야)

의사결정트리 - 분류

총 200개의 제품 데이터를 분석하여, '코팅 불량'의 핵심 원인을 찾아냅니다.



정상과 불량이 50:50으로 섞여있어, 원인을 예측하기 매우 어려운
'불순도가 높은' 상태입니다.



A 머신러닝 예제 (제조분야)

의사결정트리 - 분류



3단계: 실행 가능한 규칙(Rule)으로 변환하기

완성된 지도는 현장에서 바로 적용할 수 있는 명확한 'IF-THEN' 규칙을 알려줍니다.

🔥 규칙 1: 고위험 불량

IF 시간이 30초 미만 AND (습도가 60% 이상 OR 온도가 80°C 미만) THEN 즉각적인 공정 개선 필요!

✅ 규칙 2: 최적 정상

IF 시간이 30초 이상 AND 습도가 60% 미만 THEN 최적의 생산 조건으로 유지!



머신러닝 예제 (제조분야)

의사결정트리 - 분류

의사결정트리의 첫 질문은 어떻게 정해질까?

시작: 데이터의 불순도는?

총 200개의 제품 데이터 (정상 100개, 불량 100개)가 있습니다.



부모 노드 불순도 (Gini) = 0.5



머신러닝 예제 (제조분야)

의사결정트리 - 분류

데이터의 순도를 측정하는 법

정보 불순도(Impurity)의 정의

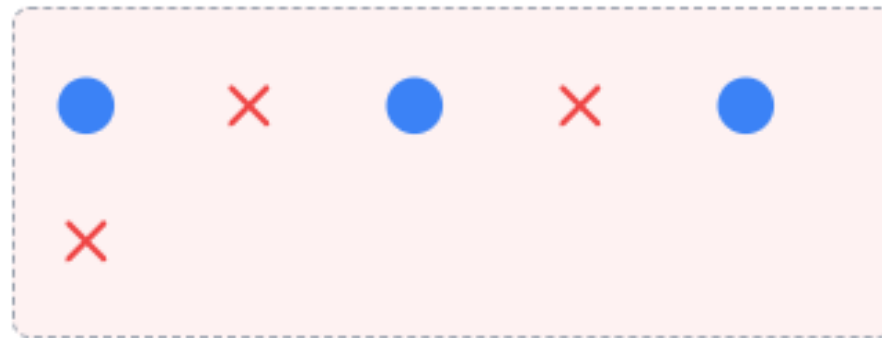
데이터의 순수함 또는 혼합된 정도를 나타내는 지표.

✓ 순도 100% (Pure Data)



하나의 분류로만 구성된 상태.

✗ 불순도 높음 (Impure Data)



여러 분류가 혼합된 상태.



머신러닝 예제 (제조분야)

의사결정트리 - 분류

데이터의 순도를 측정하는 법

불순도의 수치적 표현



정보 엔트로피

데이터의 ****무질서도**** 측정.
결과의 불확실성이 클수록 값 증가.



지니 계수

데이터의 ****불평등도**** 측정.
데이터 분포의 불평등도를 나타냄.

★ 핵심 포인트 ★

두 지표 모두 **값이 0에 가까울수록** 데이터의 순도 높음.



머신러닝 예제 (제조분야)

의사결정트리 - 분류

데이터의 순도를 측정하는 법

정보 엔트로피

$$\text{Entropy} = - \sum_{i=1}^c p_i \log_2(p_i)$$

- c : 분류 종류의 수 (예: 합격/불합격, 2개)
- p_i : 특정 분류(i)에 속할 확률

- A 생산 라인 (합격 10개, 불합격 0개)
 - 합격 확률: $p_{\text{합격}} = 10/10 = 1$
 - 불합격 확률: $p_{\text{불합격}} = 0/10 = 0$
 - 엔트로피 = $-(1 \times \log_2(1) + 0 \times \log_2(0)) = 0$
- B 생산 라인 (합격 5개, 불합격 5개)
 - 합격 확률: $p_{\text{합격}} = 5/10 = 0.5$
 - 불합격 확률: $p_{\text{불합격}} = 5/10 = 0.5$
 - 엔트로피 = $-(0.5 \times \log_2(0.5) + 0.5 \times \log_2(0.5)) = -(0.5 \times -1 + 0.5 \times -1) = 1$



머신러닝 예제 (제조분야)

의사결정트리 - 분류

데이터의 순도를 측정하는 법

지니 계수 :

$$\text{Gini Index} = 1 - \sum_{i=1}^c p_i^2$$

- c : 분류 종류의 수
- p_i : 특정 분류(i)에 속할 확률

- A 생산 라인 (합격 10개, 불합격 0개)

- 합격 확률: $p_{\text{합격}} = 1$
- 불합격 확률: $p_{\text{불합격}} = 0$
- 지니 계수 = $1 - (1^2 + 0^2) = 1 - 1 = 0$

- B 생산 라인 (합격 5개, 불합격 5개)

- 합격 확률: $p_{\text{합격}} = 0.5$
- 불합격 확률: $p_{\text{불합격}} = 0.5$
- 지니 계수 = $1 - (0.5^2 + 0.5^2) = 1 - (0.25 + 0.25) = 0.5$

A 머신러닝 예제 (제조분야)

의사결정트리 - 분류

3. 핵심 원리: 어떻게 '최적의 질문'을 선택하는가?

목표는 단 하나의 질문으로 그룹을 가장 ****순수****하게 만드는 것. 컴퓨터는 질문 후 그룹의 ****평균 불순도****가 얼마나 낮아지는지 계산합니다.

이때 **평균 불순도를 가장 낮게 만드는 질문**이 최적의 질문, 즉 ****노드의 우선순위****가 됩니다.

총 200개의 제품 데이터를 분석하여, '코팅 불량'의 핵심 원인을 찾아냅니다.



정상과 불량이 50:50으로 섞여있어, 원인을 예측하기 매우 어려운 ****불순도가 높은**** 상태입니다.

실전 예제: 어떤 질문이 최우선 순위일까?

초기 상태: 부품 200개 (정상 100, 불량 100)

초기 불순도(지니 계수) = 0.500 (최고 혼돈 상태)

질문 1: "온도 > 80°C 인가?"

'예'(100개) 그룹: 정상 75, 불량 25
'아니오'(100개) 그룹: 정상 25, 불량 75

질문 후 평균 불순도
0.375

질문 2: "습도 > 60% 인가?"

'예'(100개) 그룹: 정상 18, 불량 82
'아니오'(100개) 그룹: 정상 82, 불량 18

질문 후 평균 불순도
0.295

WINNER!

질문 3: "시간 > 30초 인가?"

'예'(100개) 그룹: 정상 92, 불량 8
'아니오'(100개) 그룹: 정상 8, 불량 92

질문 후 평균 불순도
0.147

🏆 ****결정:**** "처리 시간" 질문이 질문 후 **평균 불순도를 가장 낮게 만들었으므로**, 첫 번째 질문(뿌리 노드)으로 선택됩니다.

머신러닝 예제 (제조분야)

의사결정트리 - 분류

언제 나무는 성장을 멈추는가? (앞 마디의 조건)



조건 1: 완벽한 순도 달성

나뉘진 그룹에 한 종류(예: '정상')의 데이터만 남았을 때. 더 이상 질문이 필요 없으므로 분기를 멈춥니다.



조건 2: 정보 획득량 없음

어떤 질문을 던져도 그룹의 불순도가 더 이상 줄어들지 않을 때. 의미 없는 질문이라 판단하고 멈춥니다.



조건 3: 데이터 개수 부족

노드에 남은 데이터가 너무 적어(예: 5개 미만) 통계적 의미가 없을 때. 우연에 의한 결과를 막기 위해 멈춥니다.

머신러닝 예제 (제조분야)

의사결정트리 - 분류

✔ 장점 (Pros)



이해하기 쉬운 모델 구조

트리 구조로 결과를 시각화하여 비전문가도 쉽게 해석 가능합니다.



간편한 데이터 전처리

데이터 정규화나 스케일링 과정이 거의 필요 없습니다.



빠른 예측 속도

생성된 모델의 예측 속도가 매우 빠릅니다.

⊗ 단점 (Cons)



과적합(Overfitting) 발생 위험

훈련 데이터에 과도하게 최적화되어, 새로운 데이터에 대한 성능이 저하될 수 있습니다.



데이터에 따른 불안정성

데이터가 약간만 달라져도 트리 구조가 크게 바뀔 수 있습니다.



복잡한 경계 표현의 한계

축에 평행한 분할만 가능하며, 데이터의 경계가 복잡할 경우 예측력이 떨어집니다.

결론 및 보완책

의사결정트리는 명확한 장점이 있지만 과적합 등의 단점도 뚜렷합니다. 이러한 단점을 보완하기 위해 여러 트리를 결합하는 **랜덤 포레스트**, **그래디언트 부스팅**과 같은 앙상블 모델이 널리 사용됩니다.



머신러닝 예제 (제조분야)

랜덤포레스트 - 분류

랜덤 포레스트: 집단지성의 힘

문제: 똑똑한 나무 하나의 한계

의사결정트리는 강력하지만...

과적합: 연습문제는 잘 풀지만, 실전 시험에 약함

불안정성: 데이터가 조금만 바뀌어도 결과가 크게 흔들림



해법: 집단지성을 이용한 숲

여러 나무의 의견을 종합!

정확도 향상: 서로의 약점과 편견을 보완

안정성 확보: 더 신뢰할 수 있는 결과 도출



머신러닝 (제조분야)

랜덤포레스트 - 분류

랜덤 포레스트: 집단지성의 힘

랜덤 포레스트 생성 과정: '랜덤'의 두 가지 비밀

비밀 1: 데이터 무작위 선택 (배깅)

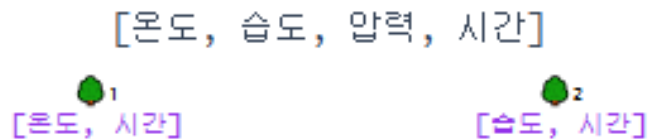
원본 데이터에서 무작위로 데이터를 뽑아
각 나무를 위한 미니 훈련 세트를 만듭니다.



모든 나무는 서로 다른 데이터를 학습

비밀 2: 특성 무작위 선택

가지를 나눌 때, 일부 특성만 무작위로 선택해
질문 후보로 사용합니다.



나무마다 다른 관점으로 문제를 보도록 강제

A 머신러닝 예제 (제조분야)

랜덤포레스트 - 분류

랜덤 포레스트:

최종 결정: 디

수백 개의 나무들이
독립적으로 예측





머신러닝 예제 (제조분야)

K-means

제조 공정 데이터 분석

K-means 군집화 심층 분석

문제 상황: 정답 없는 데이터의 홍수

실제 제조 현장에서는 '정상/불량'과 같이 명확한 정답(Label)이 없는 데이터가 대부분입니다. 수많은 센서 값, 설비 로그 속에서 어떻게 유의미한 패턴을 발견할 수 있을까요?

해결책: 비지도학습(Unsupervised Learning)의 대표 알고리즘, K-means를 통해 데이터 스스로 그룹을 형성하게 합니다.

지도학습



정답(Label) 있음
ex) 불량 예측

VS

비지도학습



정답 없음
ex) 설비 상태 군집



머신러닝 예제 (제조분야)

K-means

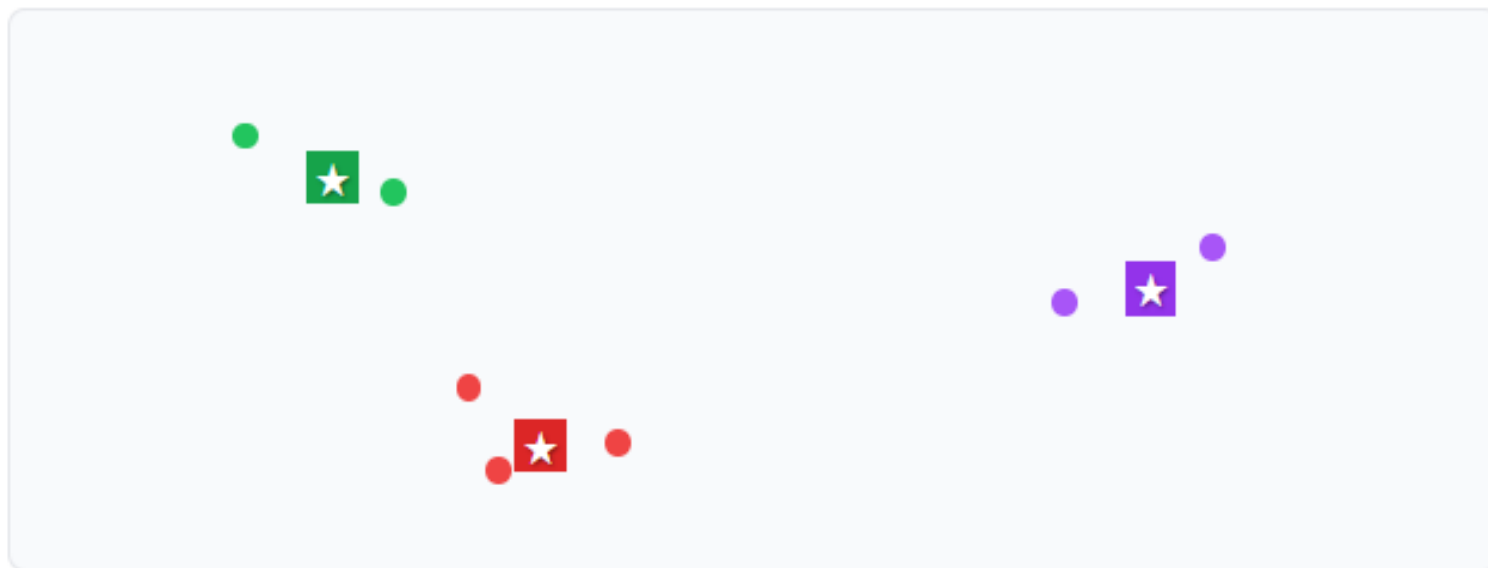
K-means 동작 원리: 상세 시각화

예시: CNC 머신의 '진동 주파수'와 '표면 온도' 데이터를 통한 설비 상태 군집화

5

수렴 및 최종 군집 확정

그룹 할당과 중심점 이동을 반복하면, 더 이상 중심점이 변하지 않는 ****수렴(Convergence)**** 상태에 도달합니다. 이 때 3개의 안정적인 최종 군집이 확정됩니다.





머신러닝 예제 (제조분야)

K-means

핵심 질문: 최적의 군집 개수(K)는 어떻게 찾을까?

엘보우 방법 (Elbow Method)

가장 널리 사용되는 방법 중 하나입니다. K값을 1부터 늘려가면서 각 K에 대한 군집 내 오차 제곱합(SSE)을 계산합니다. 그래프에서 ****기울기가 급격히 완만해지는 지점(Elbow Point)****이 최적의 K값이 될 가능성이 높습니다.

※ 이 지점부터는 K를 늘려도 군집화의 개선 효과가 크지 않다는 의미입니다.



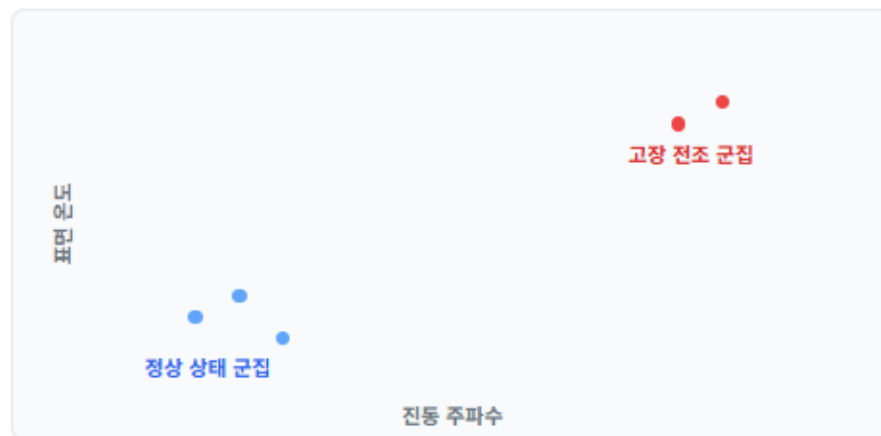
머신러닝 예제 (제조분야)

K-means

제조 현장 상세 활용 사례

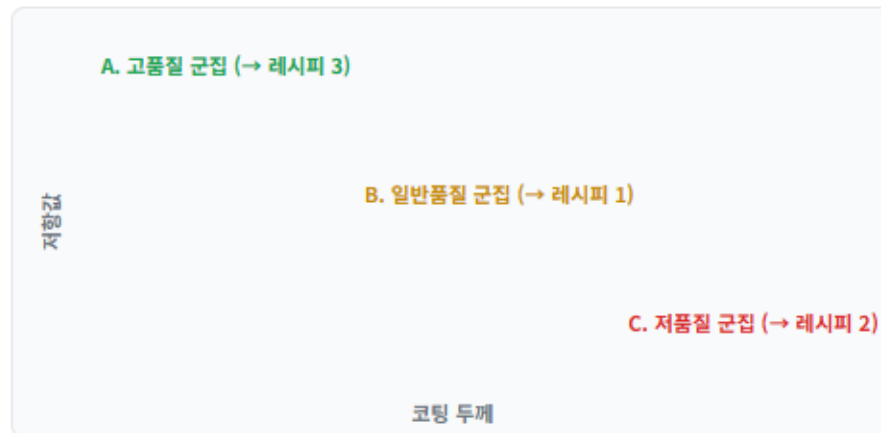
사례 1: 설비 고장 예측 (모터 베어링)

정상 상태의 설비는 센서 데이터(진동, 온도)가 하나의 큰 군집을 형성합니다. 시간이 지나면서 일부 데이터가 **고온-고진동 영역에 새로운 군집을 형성**하기 시작하면, 이는 베어링 마모 등 고장의 전조 증상으로 해석할 수 있습니다.



사례 2: 공정 레시피 최적화 (배터리 전극 코팅)

완성된 배터리 전극의 '코팅 두께'와 '저항값'을 군집화하여 3개의 품질 그룹을 발견했습니다. 각 군집이 어떤 공정 레시피(코팅 속도, 건조 온도)에서 생산되었는지 역추적하여, **'고품질 군집'을 만드는 최적의 공정 레시피를 표준화**할 수 있습니다.





제조특화 AI 플랫폼 KAMP



제조특화 AI 플랫폼 KAMP

KAMP 소개

<https://www.kamp-ai.kr/>





제조특화 AI 플랫폼 KAMP

KAMP 소개

혹시, 현장에서 이런 고민 없으신가요?



고질적인 불량

원인 모를 불량, 반복되는 문제



갑작스런 고장

예측 힘든 설비 다운타임



어려운 의사결정

경험과 직감에만 의존



제조특화 AI 플랫폼 KAMP

KAMP 소개

경험에서 데이터로, 스마트 팩토리의 시작



1단계: 데이터 수집

온도, 압력, 진동 등 설비와 센서로부터 데이터를 모읍니다.



2단계: 데이터 분석 (AI)

AI가 패턴을 발견하고 원인을 분석하여 미래를 예측합니다.



3단계: 가치 창출

품질 향상, 설비 예지보전, 공정 최적화를 이뤄냅니다.



제조특화 AI 플랫폼 KAMP

KAMP 소개

KAMP

코딩 없이 시작하는 제조 AI 플랫폼



데이터 레이크

제조 데이터, 한 곳에 안전하게!



데이터 시각화

복잡한 데이터, 한눈에 쉽게!



AI 분석 솔루션

클릭 몇 번으로 AI 모델 완성!



솔루션 스토어

검증된 AI 솔루션을 간편하게!



제조특화 AI 플랫폼 KAMP

KAMP 소개

오늘의 목표: 4시간 만에 AI 분석 전문가 되기!

1



데이터 업로드

우리 공장 데이터를 KAMP에 올리
기

2



데이터 시각화

데이터를 눈으로 보며 문제점 엿
보기

3



AI 모델 생성

클릭만으로 불량 예측 AI 만들기

GOAL! AI가 찾아낸 불량 핵심 원인 발견하기



제조특화 AI 플랫폼 KAMP

KAMP 소개

Ready to Begin?



KAMP 포털에 접속해주세요!

www.kamp-ai.kr

제조특화 AI 플랫폼 KAMP

KAMP 소개



5G + AI 스마트공장

- 수요·공급예측
지능형 가치사슬
- 제품시뮬레이션
설계 자동화
- 디지털트윈
최적공정제어
- 머신비전
품질예측
- 예지보전
설비이상예측

A

제조특화 AI 플랫폼 KAMP

KAMP 소개

설비	제목	내용	유형	구분	조회수
	제조 현장용 OCR학습 제조AI 데이터셋	충진중량을 Optical Character Recognition으로 DB에 자동기록	jpg	공정	4526
	Ford 엔진 진동 AI 데이터셋	Ford 엔진 진동 데이터를 활용한 모터 이상탐지 데이터	txt	설비	3114
	CNC 머신 AI 데이터셋	최적 가공조건 분석을 위한 CNC 머신 데이터	csv	품질	4379
	사출성형기 AI 데이터셋	사출 성형 최적 조건 분석을 위한 사출성형기 데이터	csv	품질	3595
	용접기 AI 데이터셋	용접 불량 예측 분석을 위한 로봇 용접기 데이터	csv	품질	2160
	머신비전 AI 데이터셋	열화상 이미지를 이용한 양/불량 판정을 위한 머신비전 데이터	csv	품질	2459
	프레스기 AI 데이터셋	프레스벤딩 성형 최적 조건 분석을 위한 프레스 가공 데이터	csv	품질	1752
	용해탱크 AI 데이터셋	식품 가공 공정에서의 품질 예측을 위한 원재료 공급설비(용해탱크) 데이터	csv	품질	1468
	교반구동장치 AI 데이터셋	3축(x,y,z) 진동데이터 기반 설비 예지보전을 위한 교반구동장치 데이터	csv	설비	1392
	살균기 AI 데이터셋	식품 살균 공정에서의 품질 예측을 위한 살균기 데이터	csv	품질	1691
	건조구동장치 AI 데이터셋	3축(x,y,z) 진동데이터 기반 설비 예지보전을 위한 건조 구동장치 데이터	csv	설비	1231
	X-ray 검사장비 AI 데이터셋	이미지 데이터 기반 이물질 검출을 위한 X-ray 검사장비 데이터	bmp	품질	2198

제조 현장용 OCR학습 제조AI 데이터셋

충진중량을 Optical Character Recognition으로 DB에 자동기록

제조AI데이터셋 요약

NO	구분	내용
1	분석 목적(현장이슈, 목적)	제관 과정의 무게 충진 공정으로 Count-Weight 생산품의 핵심 품질을 좌우하는 공정이다 . 충진량이 정확해야 하며 이를 이미지로 기록하고 있다 . 지속적인 관리를 위하여 이미지에서 충진중량을 광학 문자 인식 (Optical Character Recognition) 으로 DB 에 자동기록할 예정이다.
2	데이터셋 형태 및 수집방법 (csv, json, image 파일 등)	분석에 사용된 변수: Image 데이터 및 그에 수반된 Tagging XML 데이터 수집 방법: 클라우드 기반 RPMS로 현장 이미지 확보 데이터셋 파일 확장자: jpg(이미지), xml(라벨 정보)
3	데이터 개수 / 데이터셋 총량	데이터 개수 : 27,237개 데이터셋 총량 : 76.1 MB
4	분석적용 알고리즘	알고리즘 확률적 경사 하강법(stochastic gradient descent)를 활용한 딥러닝 기반의 이미지 분류 알고리즘 간략소개 라벨링된 이미지 데이터를 통해 뉴럴 네트워크를 이미지 분류 학습한다. 라벨링된 데이터로부터 손실 함수(loss function)를 최소화하기 위하여 batch data에 대해 경사 하강법(gradient descent)를 사용하였다.
5	분석결과 및 시사점	
6	데이터셋 활용(분석)을 위한 최소 요구환경	필요 SW : Anaconda, Cuda, Python 3.8 필요 패키지 : PyTorch 1.6.0, torchvision 0.7.0, Pillow 7.2.0, matplotlib 3.3.1 [CPU] -Intel(R) Core(TM) i7-7700K [GPU] GeForce GTX 1070 [RAM] 32GB

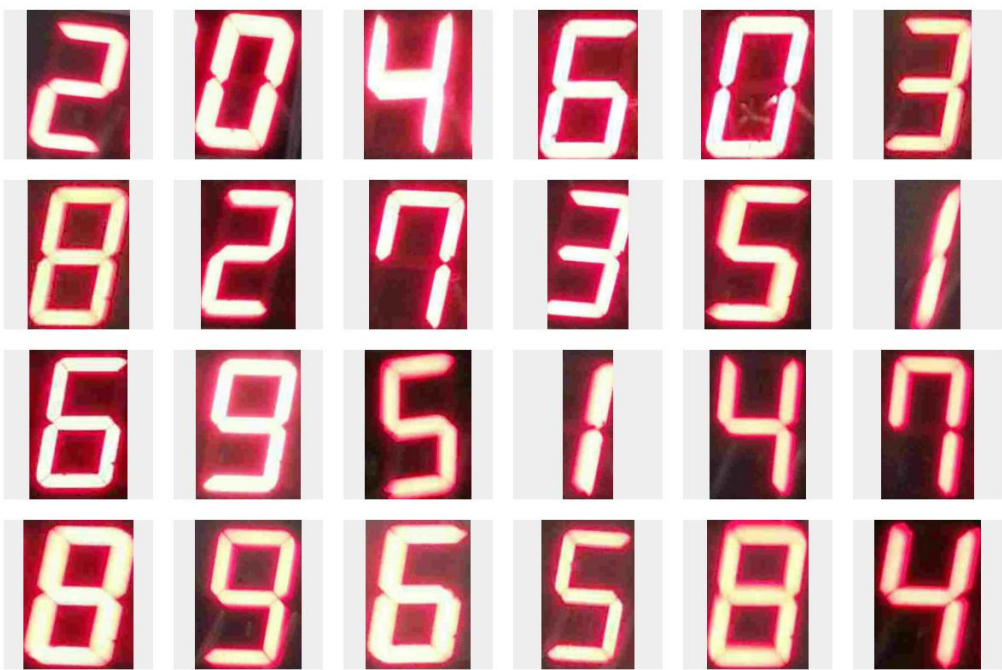
A



제조특화 AI 플랫폼 KAMP

KAMP 소개

제조AI데이터셋 형태 미리보기



분말야금 공정에 대한 데이터수집과 분석을 통한 생산 최적화

수행기간	수요기업	(주)유승
	공급기업	(주)엑센솔루션
대상공정	분말야금 금형 제품	
업종	금형LOAD_FILE_SEQ5=832	
설비	사출성형기	
문제	금형의 크기 및 공정 조건에 따라 최종 제품의 크기가 어떻게 달라지는지에 대한 추론 모델 부재	
목적	AI 솔루션을 도입 → 문제 개선 및 프레스 금형 제작 1회를 통해 의도하는 치수의 최종 제품을 얻을 수가 있게 됨	
데이터셋 형태 및 수집방법	소결 장비 전후에서 3차원 측정 시스템을 이용해 제품의 크기를 측정, 공정에 필요한 조건을 모두 제어/추론 SW와 연계, SW를 통해 데이터 회귀 분석 알고리즘을 학습, 의도된 치수를 얻을 수 있는 장비/환경 조건을 추론, 공정 및 환경에 반영.	
데이터 갯수 / 데이터셋 총량	수집대상은 총 6개(분말소재, 프레스조건, 환경/장비/소재/제품 데이터)	
분석	인공신경망을 이용한 회귀 모델 -X인자(소재 종류, 프레스 압력, 프레스 온도, 공장 온도, 공장 습도, 소결 장비 온도, 최종 제품 크기) -Y인자(프레스의 크기)	



[함안 공장 분말야금 공정과 제품군]



**KAMP사이트를 이용한
CNN 실습**

A KAMP사이트를 이용한 CNN 실습

데이터 수집하기 - Anaconda Prompt 실행하기

내 드라이브 > AI education > 산인공OJT >

▼ Gemini에게 물어보기

파일 탐색

유형 ▼

사람 ▼

수정 날짜 ▼

출처 ▼

이름 ↑

예제 : Dataset 품질 이상탐지 진단(전해탈지) AI 데이터셋

Crawling

강의자료 V7.pptx

Anaconda3-2025.06-0-Windows-x86_64.exe

C:\OJT 폴더에 저장

Anaconda Prompt

(P37) C:\Users\Waskalan>

A KAMP사이트를 이용한 CNN 실습

데이터 수집하기 - 라이브러리 설치하기


공유 보기


내 PC > 로컬 디스크 (C:) > OJT > Crawling >


이름



 `__pycache__`

docs

 download .gitignore collect_links.py

 [keywords.txt](#)

 LICENSE

 main.py README.md

requirements.txt

$$\leq \exists (C:)$$

de

nda3

no Upload

ng

oad

Anaconda Prompt

```
(base) C:\Users\maskalan>conda activate P37
```

```
(P37) C:\Users\maskalan>cd\
```

```
(P37) C:\>cd ojt\crawling
```

```
(P37) C:\#0JT\Crawling>dir
```

C 드라이브의 볼륨에는 이름이 없습니다.
볼륨 일련 번호: 6648-43E2

C:\H0JT\Crawling 디렉터리

```

2025-07-19 오후 07:25 <DIR> .
2025-07-19 오후 07:25 <DIR> ..
2025-03-25 오후 11:07 1,219 .gitignore
2025-03-25 오후 11:07 11,600 collect_links.py
2025-07-19 오후 07:25 <DIR> docs
2025-03-25 오후 11:10 <DIR> download
2025-03-25 오후 11:09 33 keywords.txt
2025-03-25 오후 11:07 11,357 LICENSE
2025-03-25 오후 11:07 15,322 main.py
2025-03-25 오후 11:07 2,171 README.md
2025-03-25 오후 11:07 56 requirements.txt
2025-03-25 오후 11:09 <DIR> __pycache__
7개 파일 41,758 바이트
5개 디렉터리 30,895,529,984 바이트 남음

```



KAMP사이트를 이용한 CNN 실습

데이터 수집하기 - 라이브러리 설치하기

```
(P37) C:\#0JT\Crawling>pip install -r requirements.txt
```

A KAMP사이트를 이용한 CNN 실습

데이터 수집하기 – 데이터 정의하기

| Crawling

공유 보기

내 PC > 로컬 디스크 (C:) > OJT > Crawling >

스크 (C:)

.T

re

nda3

to Upload

ng

oad

n

이름

__pycache__

docs

download

.gitignore

collect_links.py

keywords.txt

LICENSE

main.py

README.md

requirements.txt

keywords.txt - Windows 메모장

파일(F) 편집(E) 서식(O) 보기(V) 도움말(H)

강아지

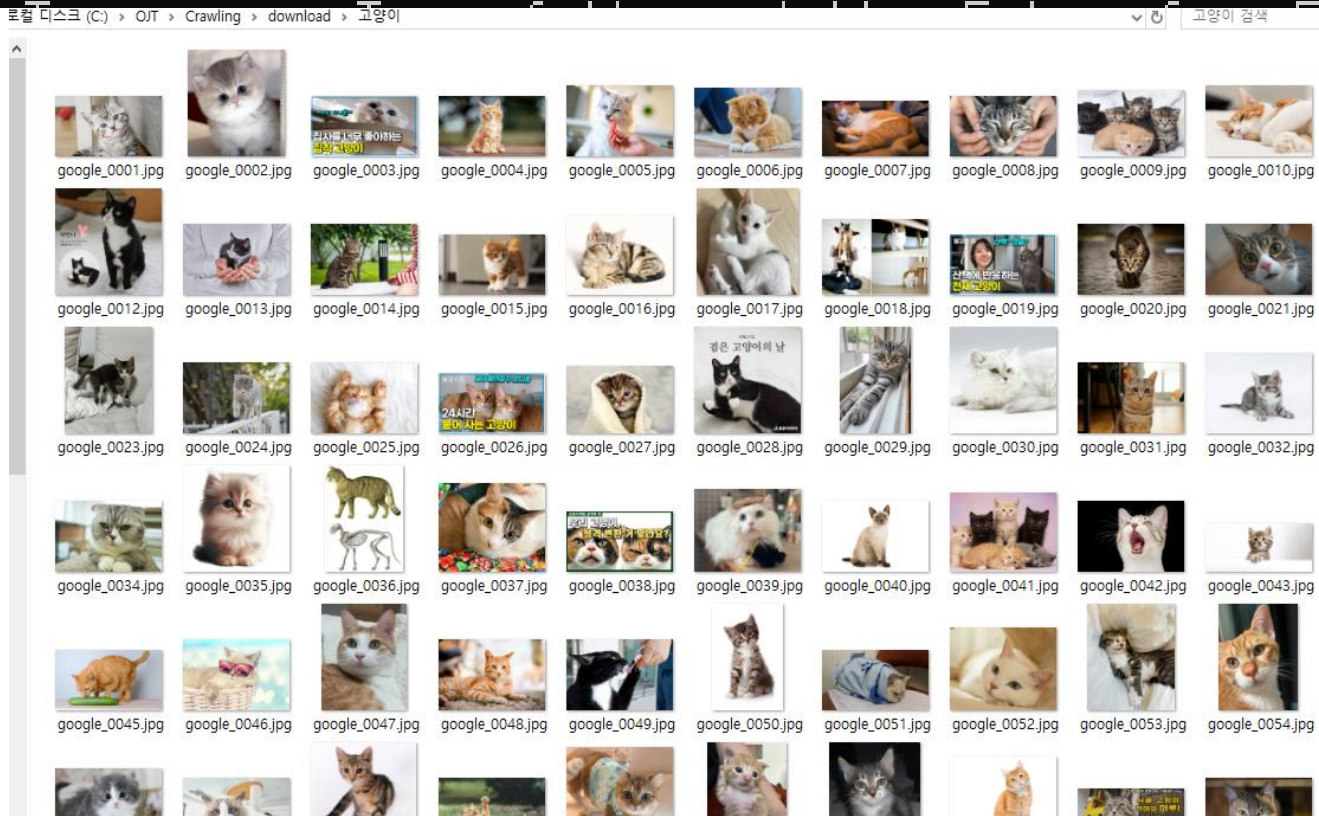
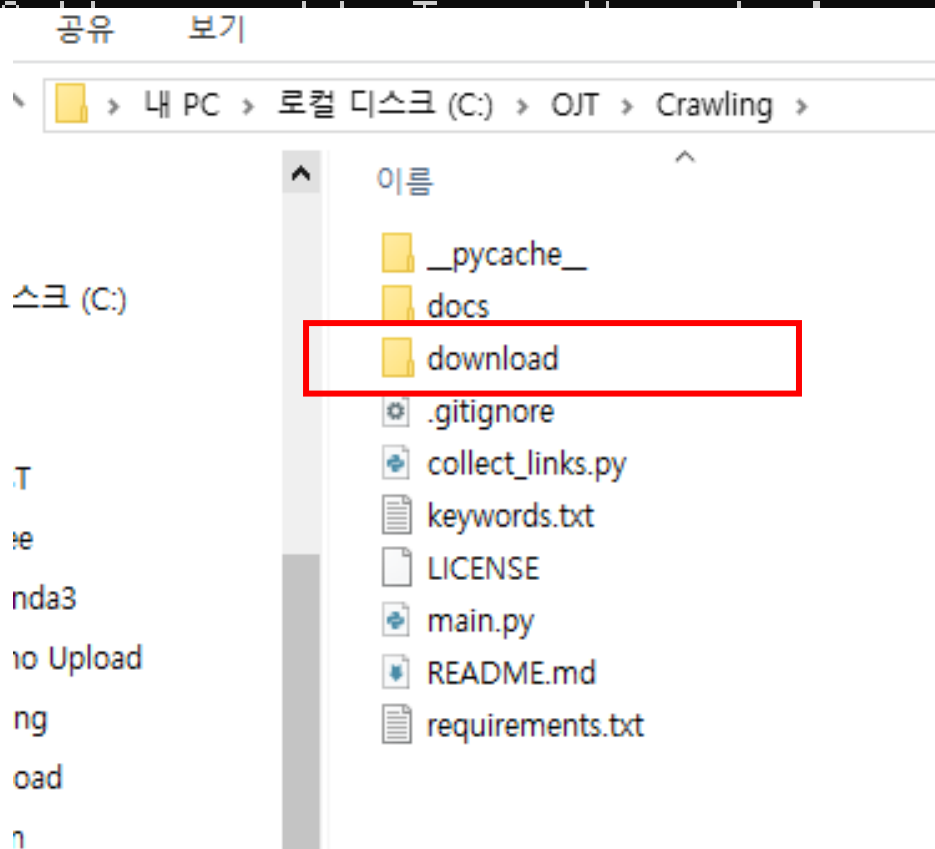
고양이

KAMP사이트를 이용한 CNN 실습

데이터 수집하기 - 데이터 수집하기

```
python main.py --skip ture --google true --naver ture --limit 50
```

```
(P37) C:\#OJT\Crawling>python main.py --skip ture --google true --naver ture --limit 50
```





KAMP사이트를 이용한 CNN 실습

데이터 수집하기 - 데이터 수집하기

로컬 디스크 (C:) > OJT > Crawling > download >

이름	수정된 날짜
강아지	2025-07
고양이	2025-07
cat.zip	2025-07
dog.zip	2025-07



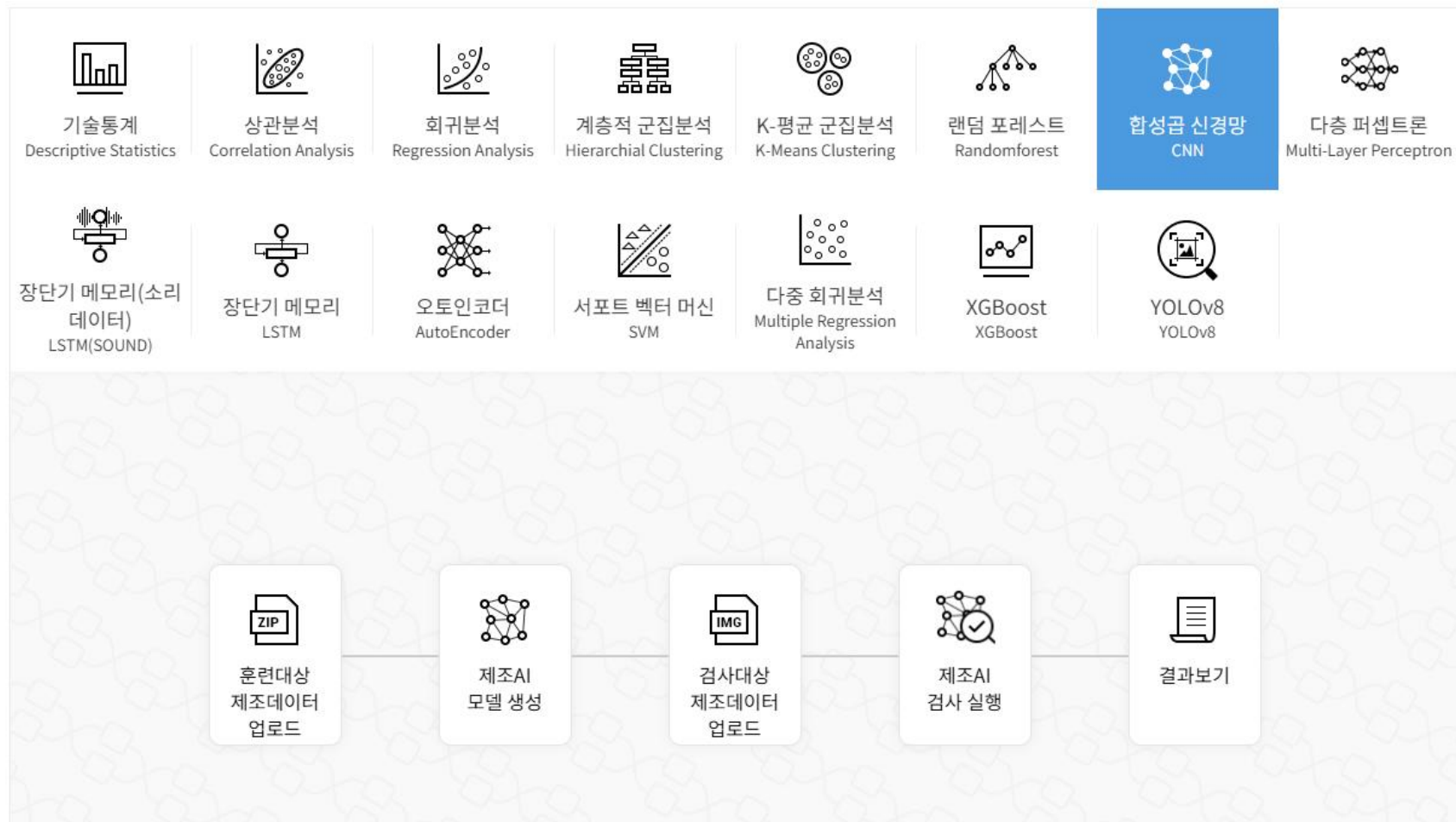
KAMP사이트를 이용한 CNN 실습

데이터 학습하기

The screenshot displays the KAMP website with a background image of industrial robotic arms. The top navigation bar includes links for '제조데이터 라이브러리', 'AI스퀘어', and '로그인'. Below this, a secondary menu highlights '제조AI분석' with a yellow underline. The main content area is divided into five columns: 'KAMP 소개' (containing 'KAMP란?' and '언론보도'), '제조AI데이터셋' (containing '무료 제조AI데이터셋', '유료 제조AI데이터셋 (제조데이터 라이브러리)', and '실증 제조AI데이터셋'), '제조AI분석' (containing '제조AI 분석도구' with a yellow underline, '제조AI 분석체험', and '제조AI 교육동영상'), '제조AI Use-Case' (containing '제조AI 적용 우수사례 (Best-Practice)', '제조AI 대동여지도', 'AI솔루션 실증사례', and '제조AI 기술 트렌드 분석'), and '알림마당' (containing '공지사항', 'KAMP 경진대회', '제조AI 솔루션 공급기업 Pool', and '이용안내'). At the bottom, a search bar contains the text '찾고 싶은 제조데이터를 검색하세요.'

A KAMP사이트를 이용한 CNN 실습

데이터 학습하기



A KAMP사이트를 이용한 CNN 실습

데이터 학습하기



훈련대상
제조데이터
업로드

제조 AI 데이터셋업로드

모델명

animal

클래스명1

dog

데이터셋1

파일첨부

dog.zip

클래스명2

cat

데이터셋2

파일첨부

고양이.zip

업로드

다음

입력 제조데이터 선택



제조AI데이터셋
선택



MY 제조데이터
업로드

닫기

A KAMP사이트를 이용한 CNN 실습

데이터 학습하기



제조AI
모델 생성

AI 모델 생성

학습모델명	animal
Epochs	20 <small>*전체 훈련데이터를 모두 학습시키는 횟수</small>
Batch Size	16 <small>* 각 Epoch마다 학습시키는 훈련데이터 수</small>
Leaning Rate	0.001 <small>* 학습될 때마다 주요 학습 변수를 갱신하는 정도</small>

AI 모델 생성

다음

00:07

AI 모델을 생성중입니다.

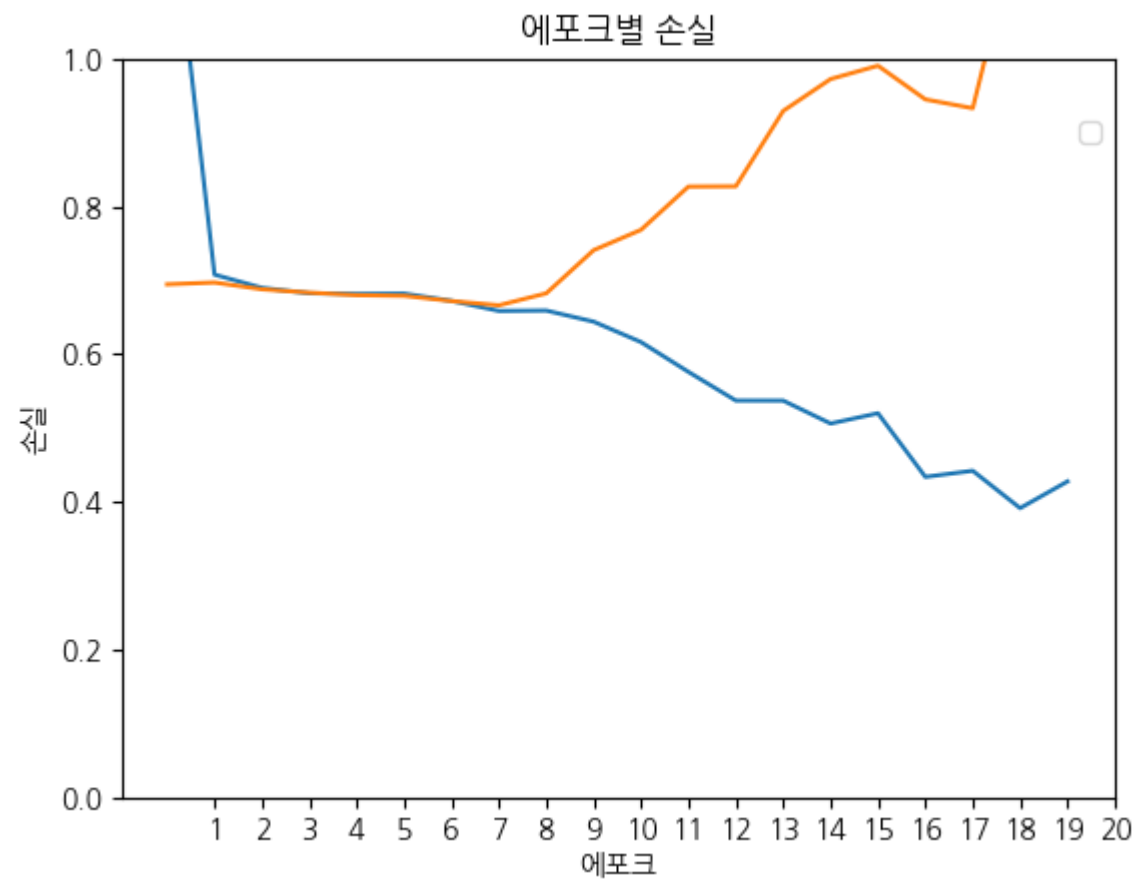
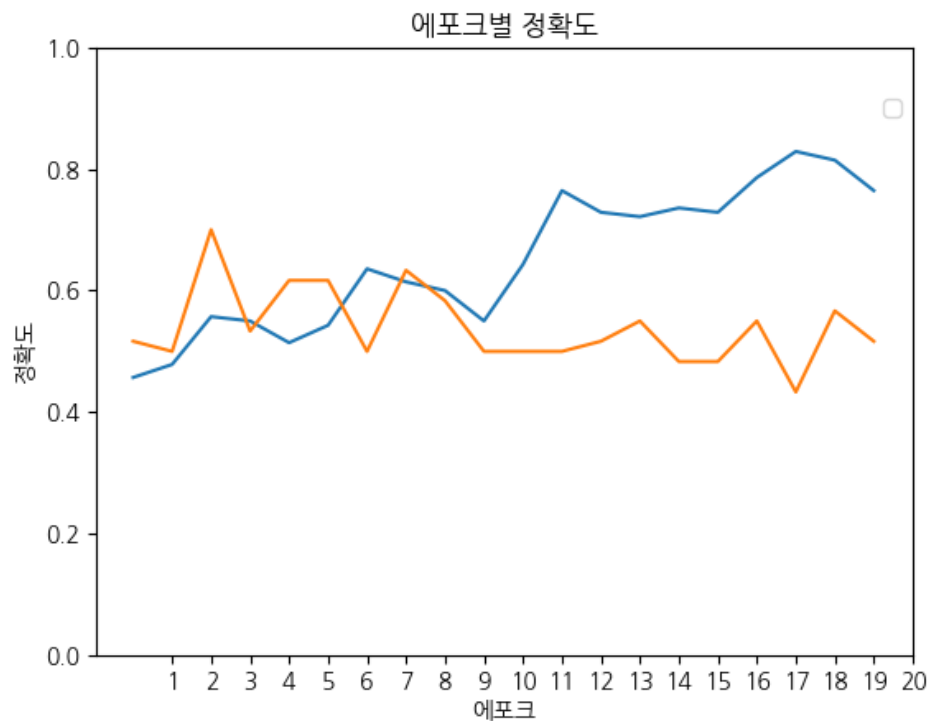
잠시만 기다려주세요.

A KAMP사이트를 이용한 CNN 실습

데이터 학습하기

2. 제조AI 모델 학습결과

1) 에포크에 따른 학습 정확도와 손실의 변화를 시각화한 그래프



A KAMP사이트를 이용한 CNN 실습

데이터 학습하기 -검사대상 업로드



검사대상
제조데이터
업로드

검사대상 제조데이터 업로드

모델명

animal

입력파일 [+] [-]

파일첨부

dog1.jpg

파일첨부

cat2.jpg

파일첨부

dog3.jpg

파일첨부

cat2.jpg

검사대상 제조데이터 업로드

검사대상 제조데이터를 업로드 하였습니다. AI 검사를 실행하여 주세요.

닫기

업로드

다음

A KAMP사이트를 이용한 CNN 실습

데이터 학습하기



제조AI
검사 실행

00:04
AI 검사 실행중입니다.
잠시만 기다려주세요.

AI 검사 실행 완료

AI 검사가 완료되었습니다. AI 검사 결과보기에서 결과를 확인하세요.

닫기

A KAMP사이트를 이용한 CNN 실습

데이터 학습하기



결과보기

합성곱신경망 검사 결과보기

dog1.jpg

c



d : 30.003%
c : 69.997%

cat1.jpg

c



d : 5.049%
c : 94.951%

dog2.jpg

d



d : 97.912%
c : 2.088%

cat2.jpg

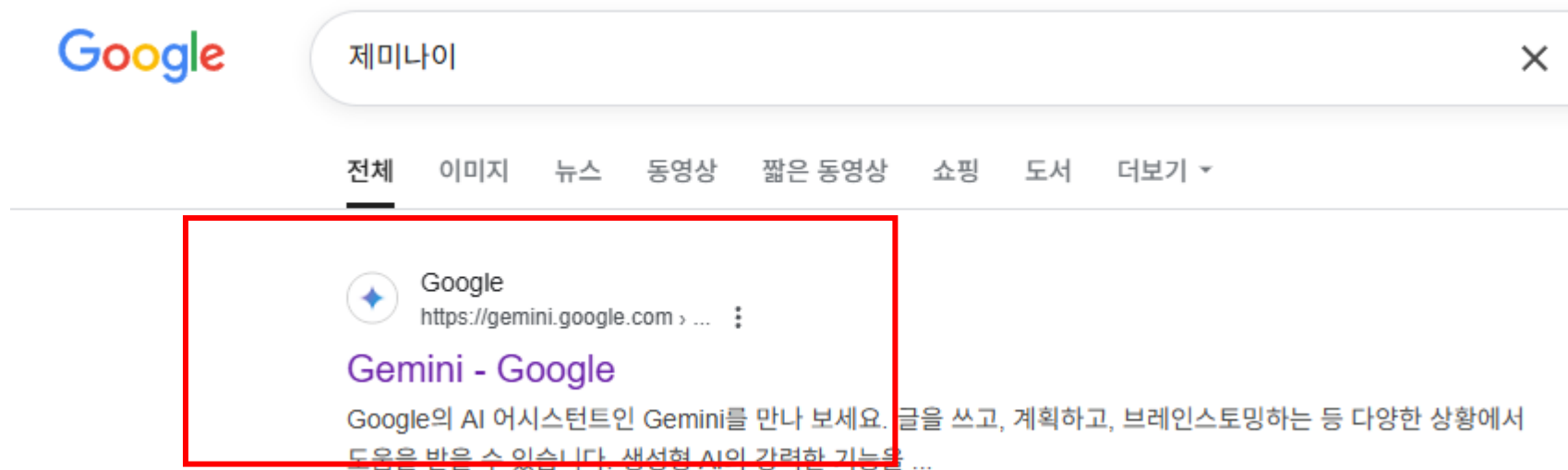
d



d : 79.998%
c : 20.002%

AI KAMP사이트를 이용한 CNN 실습

데이터 수집하기 – 데이터 수집하기





KAMP사이트를 이용한 CNN 실습

데이터 수집하기 - 데이터 수집하기

naver에서 현재 네이버 이미지 검색 결과에 맞는 정확하고 최신의 class 이름을 직접 확인하여 코드에 반영하고 WebDriverWait를 적용하여 안정성을 높여서 이미지를 크롤링하는 python 코드를 짜줘
실행하면 찾을 이미지 이름을 입력받고 다음에 찾을 이미지의 개수를 입력받도록 만들어줘.



동영상



Deep Research



Canvas



이미지



A KAMP사이트를 이용한 CNN 실습

데이터 수집하기 – 데이터 수집하기

Naver 이미지 크롤링 코드 (WebDriverWait 적용)

아래 코드는 **Selenium**과 **WebDriverWait**를 사용하여 Naver에서 특정 키워드로 이미지를 검색하고, 해당 이미지들을 지정된 폴더에 저장하는 기능을 수행합니다.

1. 필요한 라이브러리 설치

코드를 실행하기 전에 아래 라이브러리들이 설치되어 있어야 합니다. 터미널이나 명령 프롬프트에서 아래 명령어를 실행하여 설치해주세요.

Bash



```
pip install selenium beautifulsoup4 requests
```



Anaconda Prompt

```
(base) C:\Users\maskalan>conda activate P37
```

```
(P37) C:\Users\maskalan>pip install selenium beautifulsoup4 requests_
```

A KAMP사이트를 이용한 CNN 실습

데이터 수집하기 – 데이터 수집하기

내 PC > 로컬 디스크 (C:) > OJT >

이름	수정일
Crawling	2025-
Naver	2025-

2. 웹 드라이버 준비

이 코드는 크롬(Chrome) 브라우저를 기준으로 작성되었습니다. 사용하시는 크롬 브라우저 버전에 맞는 [ChromeDriver](#)를 다운로드하여 코드 파일과 같은 폴더에 두거나, 시스템 경로에 추가해야 합니다.


3. 전체 코드

Python



A KAMP사이트를 이용한 CNN 실습

데이터 수집하기 – 데이터 수집하기

 chrome for developers

아이디어 얻기 블로그 Docs ▾ Chrome의 새로운 기능

ChromeDriver

개요

Capabilities 및 Chrome Options

Chrome 확장 프로그램

참여

디자인 문서

ChromeDriver 다운로드

Chrome 버전 115 이상

이전 버전의 Chrome

안정화 버전

카나리아 릴리스

버전 선택

translated by Google 이 페이지는 [Cloud Translation API](#)를 통해 번역되었습니다.

홈 > Docs > ChromeDriver

다운로드

⚠ 경고: Chrome 버전 115 이상을 사용하는 경우 [테스트용 Chrome 사용 가능 여부 대시보드](#)를 참고하세요
버전을 다운로드할 수 있는 [JSON 엔드포인트](#)가 있습니다.

이전 Chrome 버전

이전 버전의 Chrome을 지원하는 ChromeDriver 버전은 다음과 같습니다.

적절한 버전의 ChromeDriver를 선택하는 방법에 관한 자세한 내용은 [버전 선택](#) 페이지를 참고하

ChromeDriver 114.0.5735.90

A KAMP사이트를 이용한 CNN 실습

데이터 수집하기 – 데이터 수집하기

Stable

Version: 138.0.7204.157 (r1465706)



Binary	Platform	URL	HTTP status
chrome	linux64	https://storage.googleapis.com/chrome-for-testing-public/138.0.7204.157/linux64/chrome-linux64.zip	200
chrome	mac-arm64	https://storage.googleapis.com/chrome-for-testing-public/138.0.7204.157/mac-arm64/chrome-mac-arm64.zip	200
chrome	mac-x64	https://storage.googleapis.com/chrome-for-testing-public/138.0.7204.157/mac-x64/chrome-mac-x64.zip	200
chrome	win32	https://storage.googleapis.com/chrome-for-testing-public/138.0.7204.157/win32/chrome-win32.zip	200
chrome	win64	https://storage.googleapis.com/chrome-for-testing-public/138.0.7204.157/win64/chrome-win64.zip	200
chromedriver	linux64	https://storage.googleapis.com/chrome-for-testing-public/138.0.7204.157/linux64/chromedriver-linux64.zip	200
chromedriver	mac-arm64	https://storage.googleapis.com/chrome-for-testing-public/138.0.7204.157/mac-arm64/chromedriver-mac-arm64.zip	200
chromedriver	mac-x64	https://storage.googleapis.com/chrome-for-testing-public/138.0.7204.157/mac-x64/chromedriver-mac-x64.zip	200
chromedriver	win32	https://storage.googleapis.com/chrome-for-testing-public/138.0.7204.157/win32/chromedriver-win32.zip	200
chromedriver	win64	https://storage.googleapis.com/chrome-for-testing-public/138.0.7204.157/win64/chromedriver-win64.zip	200
chrome-headless-shell	linux64	https://storage.googleapis.com/chrome-for-testing-public/138.0.7204.157/linux64/chrome-headless-shell-linux64.zip	200
chrome-headless-shell	mac-arm64	https://storage.googleapis.com/chrome-for-testing-public/138.0.7204.157/mac-arm64/chrome-headless-shell-mac-arm64.zip	200
chrome-headless-shell	mac-x64	https://storage.googleapis.com/chrome-for-testing-public/138.0.7204.157/mac-x64/chrome-headless-shell-mac-x64.zip	200
chrome-headless-shell	win32	https://storage.googleapis.com/chrome-for-testing-public/138.0.7204.157/win32/chrome-headless-shell-win32.zip	200
chrome-headless-shell	win64	https://storage.googleapis.com/chrome-for-testing-public/138.0.7204.157/win64/chrome-headless-shell-win64.zip	200

A KAMP사이트를 이용한 CNN 실습

데이터 수집하기 – 데이터 수집하기

- ✓ 오늘 (2)
- chromedriver-win64 (1)
 - chromedriver-win64 (1).zip

압축 풀어서

.드 > chromedriver-win64 (1) > chromedriver-win64

이름	수정한 날짜
chromedriver.exe	2025-07-19
LICENSE.chromedriver	2025-07-19
THIRD_PARTY_NOTICES.chromedriver	2025-07-19

내 PC > 로컬 디스크 (C:) > OJT >

이름	수정한 날짜
Crawling	2025-
Naver	2025-

d

A KAMP사이트를 이용한 CNN 실습

데이터 수집하기 – 데이터 수집하기

3. 전체 코드

Python



```
import os
import time
import requests
from selenium import webdriver
from selenium.webdriver.common.by import By
from selenium.webdriver.support.ui import WebDriverWait
from selenium.webdriver.support import expected_conditions as EC
from selenium.webdriver.common.keys import Keys

# --- 설정 ---
SEARCH_KEYWORD = "강아지" # 검색할 키워드
SAVE_DIRECTORY = "downloaded_images" # 이미지를 저장할 폴더 이름
SCROLL_PAUSE_TIME = 2 # 스크롤 후 대기 시간 (초)
```

메모장 열고

붙여넣기

이름은 naver.py로

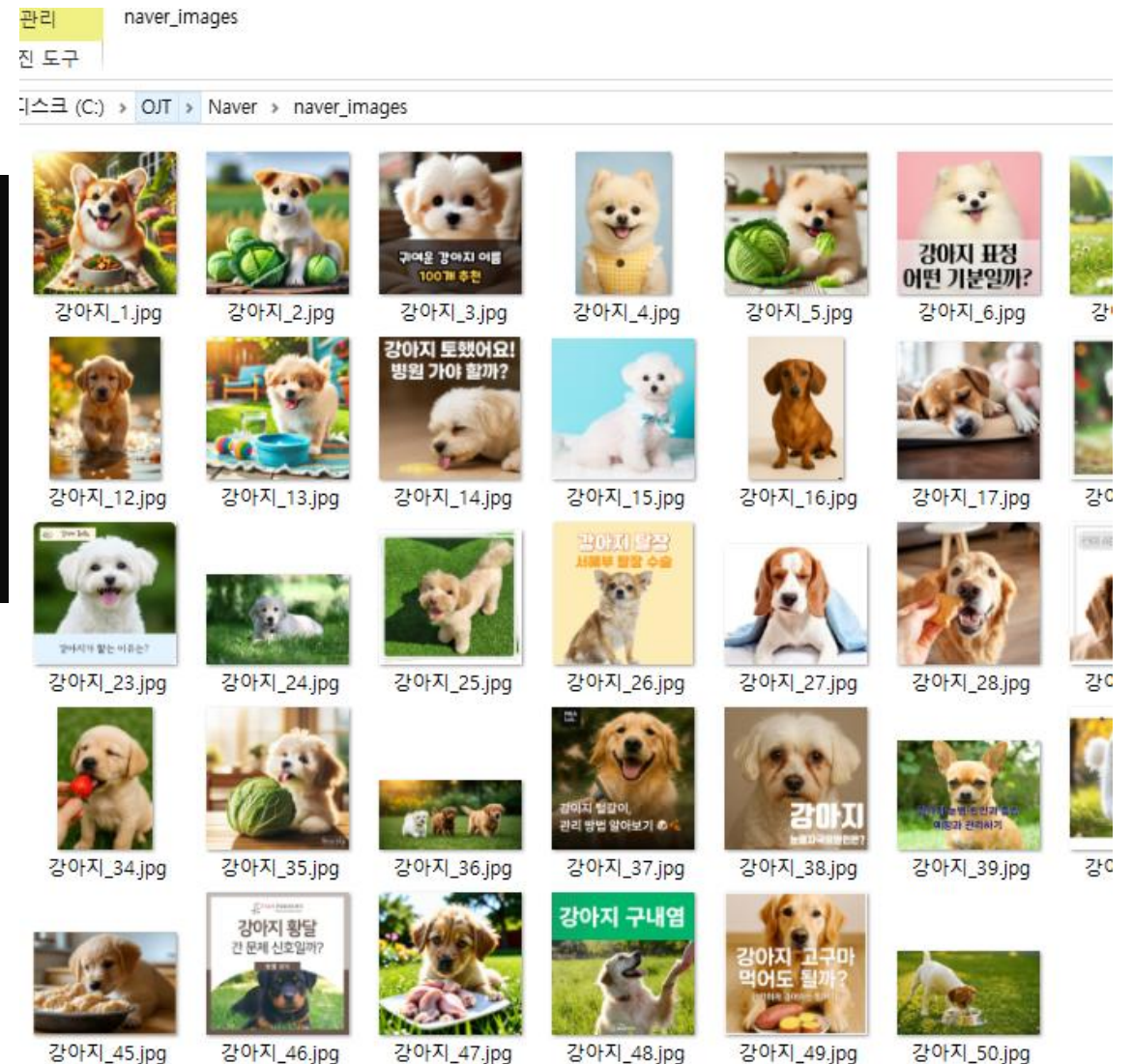
Naver폴더에 저장

A KAMP사이트를 이용한 CNN 실습

데이터 수집하기 – 데이터 수집하기

Anaconda Prompt

```
(base) C:\Users\maskalan>conda activate P37  
(P37) C:\Users\maskalan>cd\\  
(P37) C:\>cd ojt\\naver  
(P37) C:\OJT\\Naver>python naver.py_
```





프롬프트 작성을 위한 도구 활용

NoteBook LM

노트북 LM

<문서를 전문가처럼 생성하기 위한 프롬프트 작성기>



노트북LM



notebooklm.google

<https://notebooklm.google>

Google NotebookLM | AI Research Tool & Thinking Partner

Meet **NotebookLM**, the AI research tool and thinking partner that can analyze your sources, turn complexity into clarity and transform your content.

무엇이든 이해
할 수 있습니다

최신 Gemini 모델로 제작된 조사 및 사고 파트너로, 신뢰할 수 있는 정보를 기반으로 그라운드되었습니다.

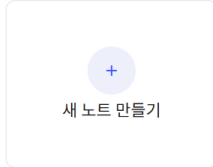
Try NotebookLM

NoteBook LM

노트북 LM

<필요한 문서를 잘 작성하기 위한 리소스를 검색>

최근 노트북



NotebookLM

소스 추가

소스를 추가하면 NotebookLM이 가장 중요한 정보에 따라 응답을 제공합니다.
(예: 마케팅 계획, 수업 자료, 연구 노트, 회의 스크립트, 판매 문서 등)

소스 검색



소스 업로드

업로드할 파일을 선택하거나 드래그 앤 드롭하세요.

지원되는 파일 형식: PDF, .txt, Markdown, 오디오(예: mp3)

Google Drive

Google Docs

Google Slides

링크

웹사이트

YouTube

텍스트 붙여넣기

복사된 텍스트

소스 한도

0/300

NoteBook LM

노트북 LM

<검색된 리소스를 참고 자료로 활용하기 위한 저장>

소스 검색

×



무엇에 관심이 있으신가요?

전라북도 한정식의 정의와 특징을 설명해줘

🔍 I'm feeling curious

제출

+ 추가

🔍 탐색

모든 소스 선택



📄 "1000년 역사, 전주한정식의 위상을 지키자" - 전주일보



📄 [외식의窓] 광주식 백반과 전주식 백반의 차이



소스 검색

×



무엇에 관심이 있으신가요?

"전라북도 한정식이 다른 지역 한정식과 어떤 점이 다른지 정리해줘."

🔍 I'm feeling curious

제출

NoteBook LM

노트북 LM

<참조한 소스를 활용하여 요청한 사항에 대한 프롬프트 작성>

AI, 학습 및 업무 효율성 혁신

출처

+ 추가

🔍 탐색

모든 소스 선택

✓

AI를 활용하여 10분만에 깔끔한 발표자료 준비하는 법 | 파워포...

✓

AI를 활용하여 논문을 효율적으로 읽고 쓰는 5가지 방법 - 지피...

✓

[강남인강] AI가 도와주는 중학 영어 수행평가 (2강)

✓

[중학교 수행평가 예시] 환경신문 만들기 과제 - 주제 추천부터...

✓

고등학교 수행평가 보고서 예시 공개! 서울대 합격생의 세특은...

✓

디지털세상 수기공모전 '디지털세상 어때요?' 디지털씨앗상 ...

✓

부정행위 사례 및 대응방안 - 성균관대학교 챗GPT 종합 안내

✓

수행평가 보고서 작성 - 독스 AI 학습도구

✓


수행평가 실험 보고서 주제 추천 - 독스 AI 학습도구

✓

코딩 없이 AI 만들기: 비개발자를 위한 쉬운 인공지능 활용법 - ...

✓

채팅



AI, 학습 및 업무 효율성 혁신

소스 10개

제공된 자료들은 인공지능(AI)의 다양한 활용 사례와 교육 환경에서의 영향을 다룹니다. 특히, 개인이 코딩 지식 없이 AI 도구를 활용하여 프레젠테이션 자료나 그림을 만드는 방법을 설명하고, 논문 작성 및 연구 과정에서 AI를 활용해 효율성을 높이는 구체적인 방법들을 제시합니다. 또한, AI가 중고등학생들의 수행평가 준비를 돕는 도구로 사용될 수 있음을 보여주지만, 동시에 AI를 활용한 부정행위의 위험성과 그에 대한 교육 기관의 대응 방안도 함께 다루며 기술의 윤리적 사용을 강조합니다.

📌 메모에 저장

📄 메모 추가

🔊 AI 오디오 오버뷰

🗺️ 마인드맵

입력을 시작하세요...

소스 10개

▶

AI 도구가 개인의 학습과 업무 효율성을 어떻게 혁신하고 있는가?

AI를 활용한 학습 및 업무 도구 사용의 주요 이점과 한계점은 무엇인

>

스튜디오

AI 오디오 오버뷰

더 많은 언어로 AI 오디오 오버뷰를 만드세요. 자세히 알아보기

🗣️ 심층 분석 대화

호스트 2명

맞춤설정

생성

노트

+ 메모 추가

📖 학습 가이드

📄 브리핑 문서

📖 FAQ

📄 타임라인

🔗 AI 활용법과 오용 사례 및 대응 방안

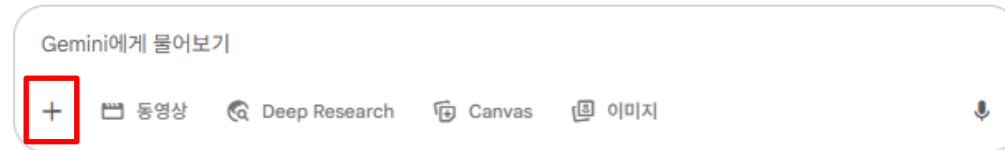
💡 클릭하여 마인드맵 열기

🔗 AI 활용법과 오용 사례 및 대응 방안

💡 클릭하여 마인드맵 열기

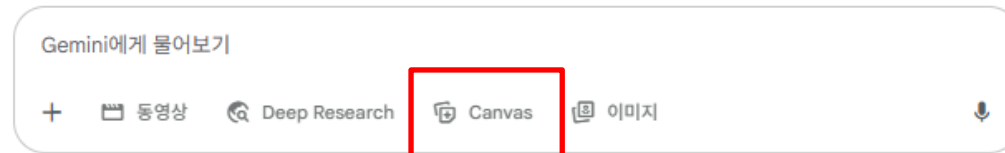
Gemini 사용

첨부자료 활용 <논문 양식을 첨부 파일로 업로드>



NoteBook LM으로 작성한 프롬프트를 입력하여 제미나이 또는 파일
첨부가 가능한 생성형 AI를 활용하여 작성

논문에 사용할 그림이나 차트, 표도 AI로 작성 가능



인포그래픽(차트나 그래프) 제작에 용이, 슬라이드나 발표자료 제작에 활용