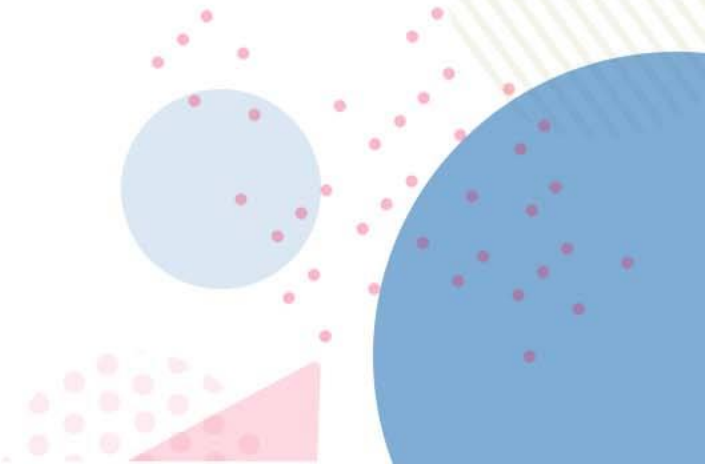


날씨 데이터를 활용한 태양광 발전량 예측

0강. 강의 소개





박

희

진

▶ 주요 강의 분야: 머신러닝, 딥러닝, 리눅스, IoT

▶ 강의 경력 (2018년 ~ 현재)

비트교육센터

서울 ICT 이노베이션 스퀘어

동북권 ICT 이노베이션 스퀘어





태양광 발전량 예측

- 필요한 기반 지식
- 데이터의 수집과 선택
- 전처리
- 딥러닝 모델링
- 모델의 학습과 예측

➡ 데이터 중심 AI 란?



날씨 데이터를 활용한 태양광 발전량 예측

1강. 데이터중심 시의 이해



1. 모델 중심 AI에서 데이터 중심 AI로
2. 과거 인공지능 논문에서 볼 수 있는 현상들
3. 모델 중심 AI VS 데이터 중심 AI
4. AI 모델 성능을 효율적으로 높이는 방법
5. 데이터 양과 질의 상관관계
6. Clean VS Noisy data
7. 인공지능과 데이터 사이언스



2021년에 앤드류 응 교수가 DeepLearning AI에서 [From Model-centric to Data-centric AI](#) 라는 강의에서 우리가 인공지능 모델을 다루는데 있어 쉽게 놓칠 수 있는 부분을 설명합니다.

우리가 다루는 인공지능은 아래와 같이 **코드**와 **데이터**로 이루어져 있습니다.

$$\text{AI System} = \text{Code} + \text{Data}$$

(모델/알고리즘)



앤드류 응 교수는

인공지능 연구의 Abstract만 보면 99%의 연구는 모델 개선에 대해 다루고,
1%의 연구는 데이터 증식에 대해 다루고 있다고 이야기 합니다.

그러면 왜 이렇게 편향적인 연구가 이루어지고 있을까요?

AI System = Code + Data
(모델/알고리즘) **표준 데이터셋**

그 이유는, 표준 데이터셋에 있습니다.



모델 중심 AI

- 데이터는 일반적인 전처리 과정만 진행하고, 모델만 반복적으로 향상하는데 집중합니다.
- 데이터는 최대한 모을 수 있는 만큼 모으고, 데이터에 노이즈가 있더라도 문제 없을 정도로만 전처리 합니다.

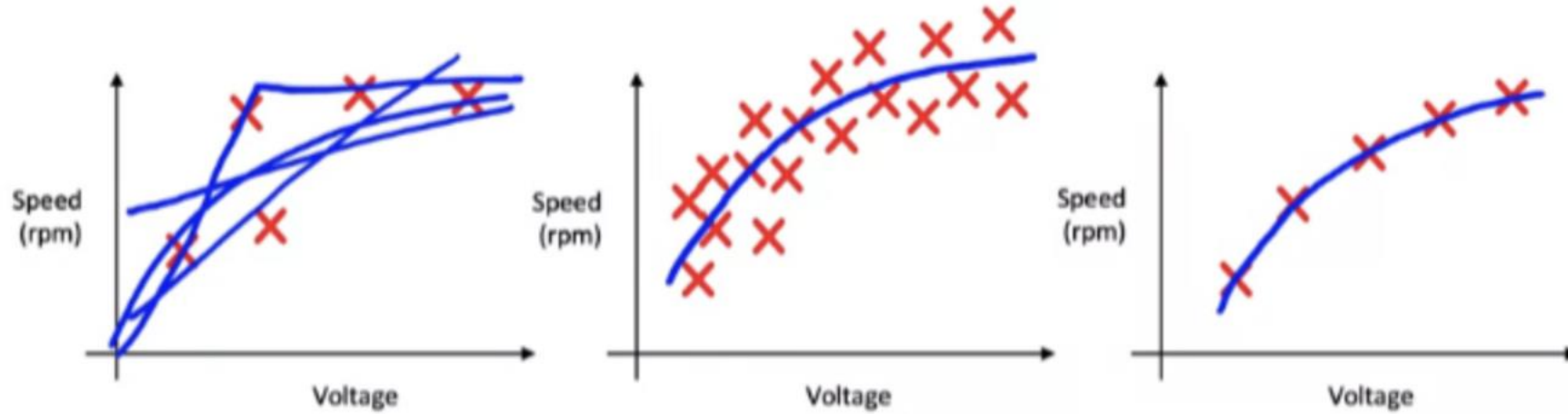
데이터 중심 AI

- 모델은 적절한 수준으로 고정하고, 데이터의 질을 우선 반복적으로 향상하는데 집중합니다.
- 데이터는 일관성 있는 레이블을 유지하는 것이 중요합니다.(레이블링 품질의 중요성)



| | Steel defect detection | Solar panel | Surface inspection |
|---------------|------------------------|--------------------|--------------------|
| Baseline | 76.2% | 75.68% | 85.05% |
| Model-centric | +0% (76.2%) | +0.04% (75.72%) | 0% (85.05%) |
| Data-centric | +16.9% (93.1%) | +3.06% (78.74%) | +0.4% (85.45%) |





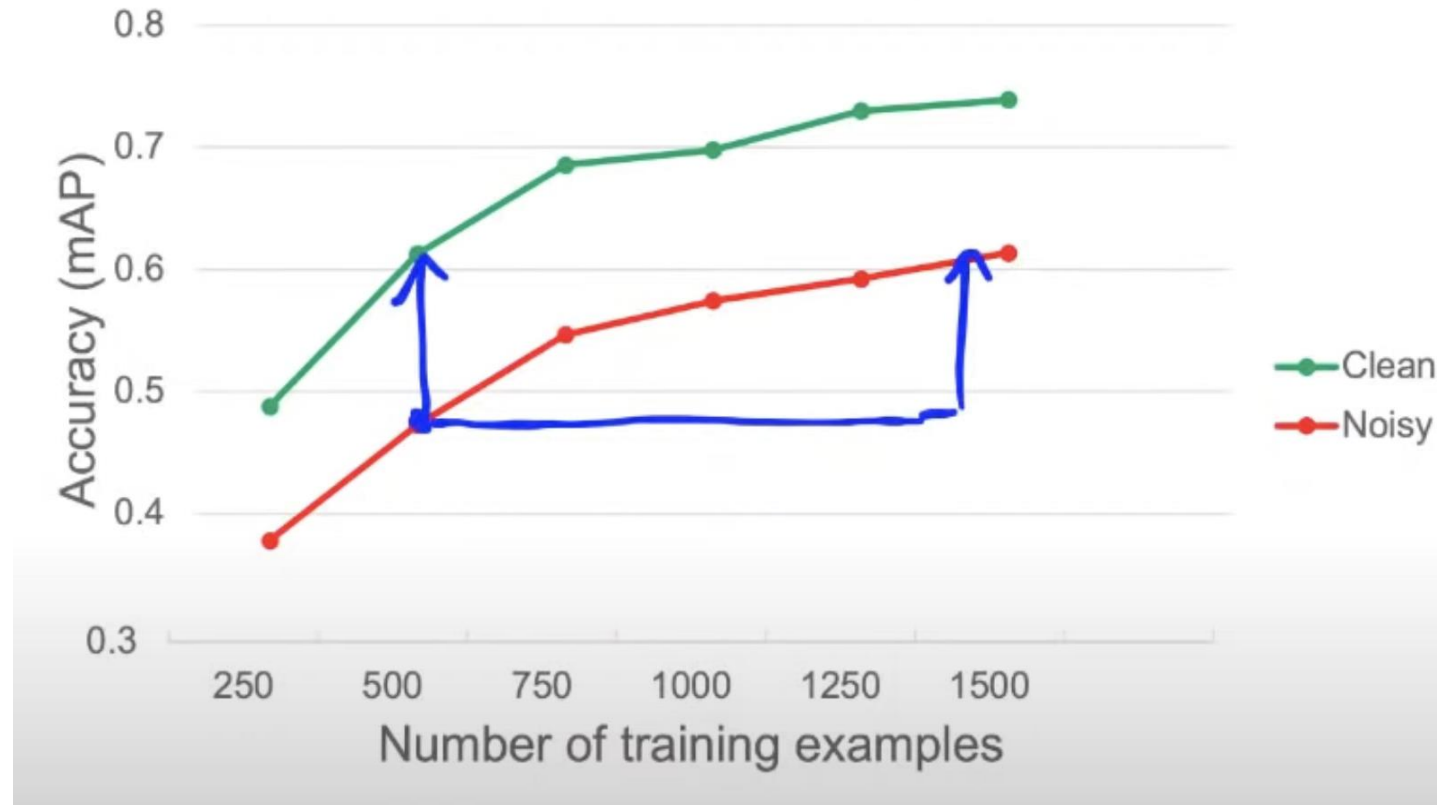
데이터가 적고,
레이블에도 노이즈가
있는 경우

데이터는 많으나,
레이블에 노이즈가
있는 경우

데이터는 적으나,
레이블에 노이즈가
없는 경우



6 Clean VS Noisy data



인공지능

지금까지 풀기 어려운 문제를
풀어낼 수 있는 방법론

단어만 가지고는 어떻게 문제를
해결하는지 이해하기 어렵다.

데이터 사이언스

양질의 데이터를 가지고
문제를 풀어내는 방법론

이미 단어에서 데이터를 기반으로
문제를 해결함을 명확하게 나타낸다.



수고하셨습니다

