

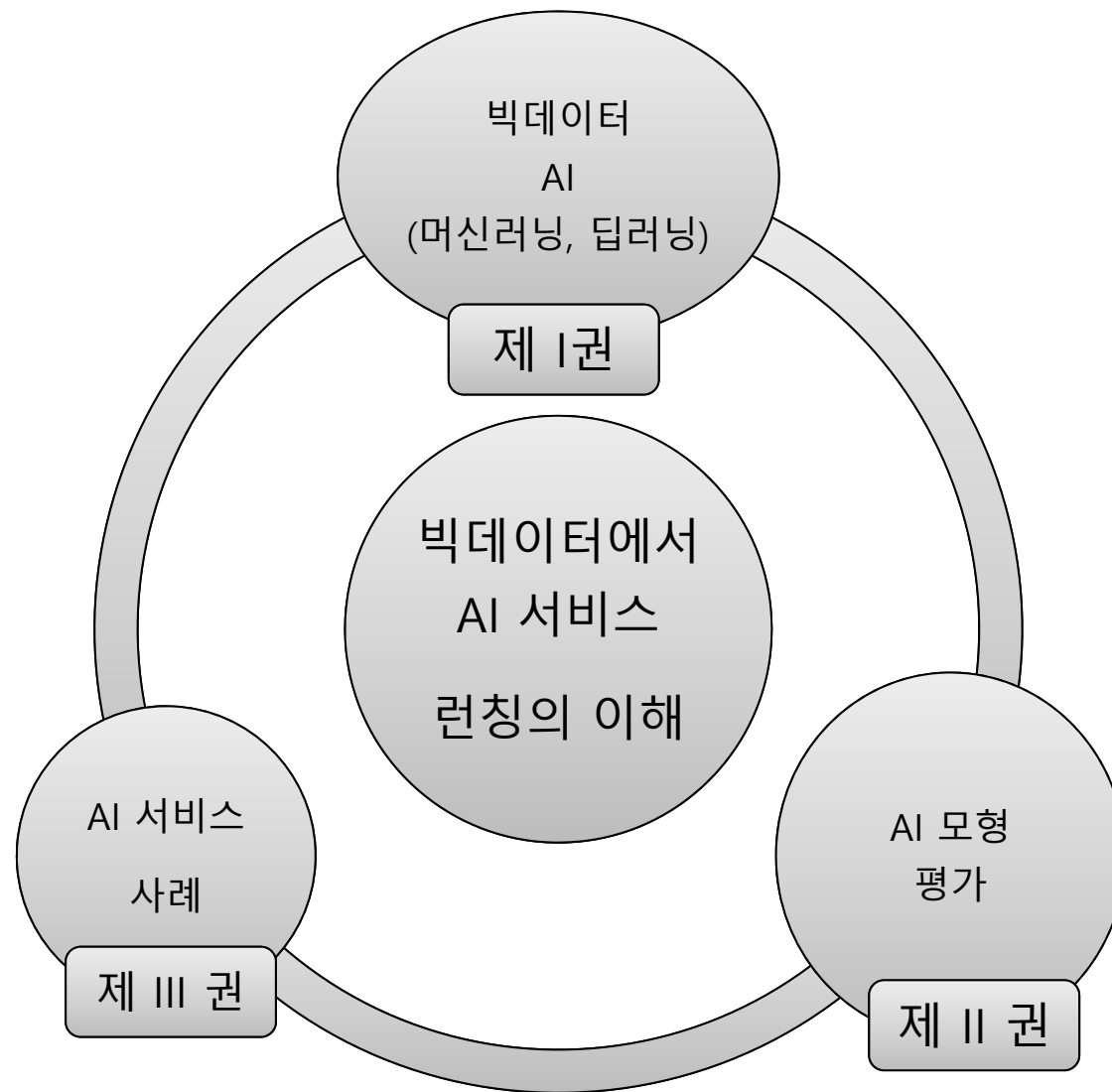
빅데이터에서 AI 서비스 모델 개발

Book Link

<http://digital.kyobobook.co.kr/digital/ebook/ebookDetail.in?selectedLargeCategory=001&barcode=480D190812700&orderClick=LAG&Kc=>

Content Table

1. 데이터의 중요성
2. Big Data 출현 배경
3. Data 측정 수준 이해
4. Big Data 란
5. 데이터에서 AI 서비스로의 단계
6. 빅데이터 기반 AI 서비스의 구성도 그리기
7. 인공지능, 머신러닝, 딥러닝
8. 목적치 유무에 따른 분류: Target 변수에 따른 머신러닝의 분류



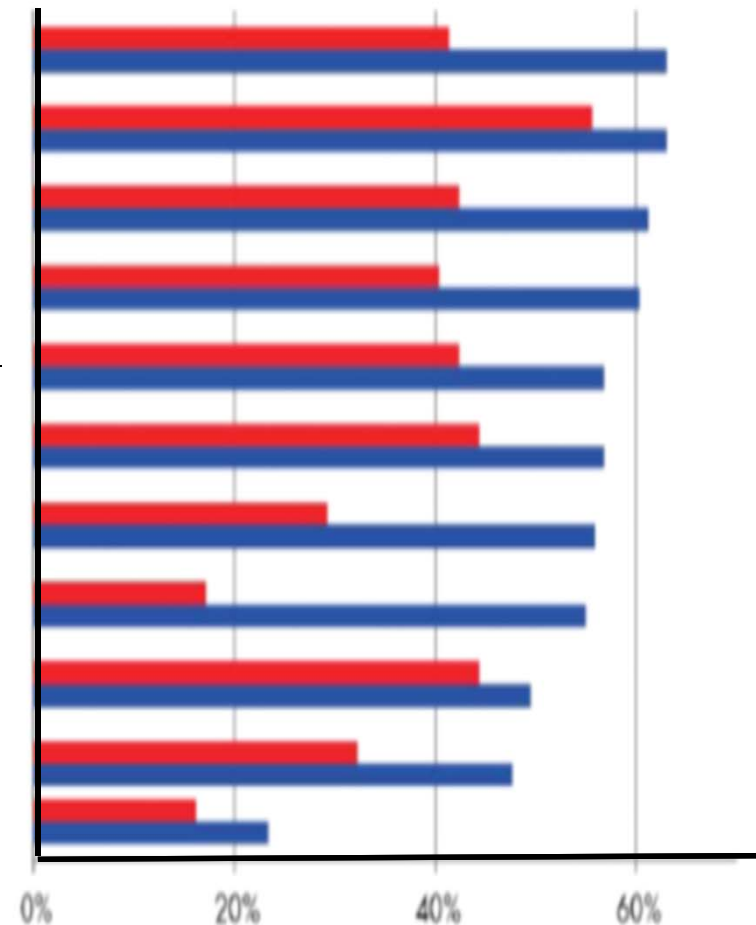
데이터의 중요성

데이터에 대한 태도 및 관점

문장에 대해 “완전히 동의합니다”인 9또는 10을 선택한 응답자의 비율

데이터를 활용하면 비즈니스 민첩성이 향상됨
효과적인 데이터 전략은 회사의 경쟁 우위가 될 수 있음
데이터 관리 전략을 개선하면 고객에게 보다 신속하게 대응할 수 있음
데이터를 전략적 자산으로 관리하고 있음
수많은 데이터 소스를 통합하는 기능을 개선하면 비즈니스의 예측 능력이 개선됨
비즈니스 데이터를 마스터링하면 기업 전체의 현업 협업이 개선됨
우리의 데이터 전략은 비즈니스 부문에 집중됨
주기적으로 비즈니스 리더와 데이터 관리 전략에 대해 의논함
데이터를 활용하여 내부 사용자의 능력을 향상시킬 수 있음
우리의 우선 순위는 비즈니스 프로세스에 데이터를 활용하는 것
우리의 데이터 관리가 현재 요구 사항을 충족하기에 충분함

■ 비임원 ■ 임원



Economist "The Data Directive"

Big Data 출현 배경

- 우리가 움직이는 지금도 기록으로 남아 데이터화 되고 있다. 전철 이동에 사용하는 카드 사용 이력, GPS 살에 기록 되는 이동 정보, 은행이나 쇼핑몰 구매와 뷰 정보등이 모두 데이터가 되고 있다.

Data 측정 수준 이해

- **(1) 변수란**
 - ① 변수는 연구의 대상이 되는 성격 또는 속성이다.
 - ② 변수는 변하는 수이다.
- **(2) 특성의 상태를 표시하는 방법, 척도에 따른 분류**
 - - 계량치(Continuous Data)는 셀 수 없는 형태로 측정된, 길이, 무게, 온도, 시간, 생산량이 된다.
 - - 계수치(Discrete Data)는 셀수 있는 형태로 측정된, 부적합품수, 부적합수, 생산량이 된다.
- **(3) 양적변수와 질적변수**
 - 양적변수(quantitative, continuous)는 키, 연봉 등과 같이 수치화된 값을 가지는 변수이고, 질적변수(qualitative, categorical)는 성별, 지역 등과 같이 코드화된 값이다.

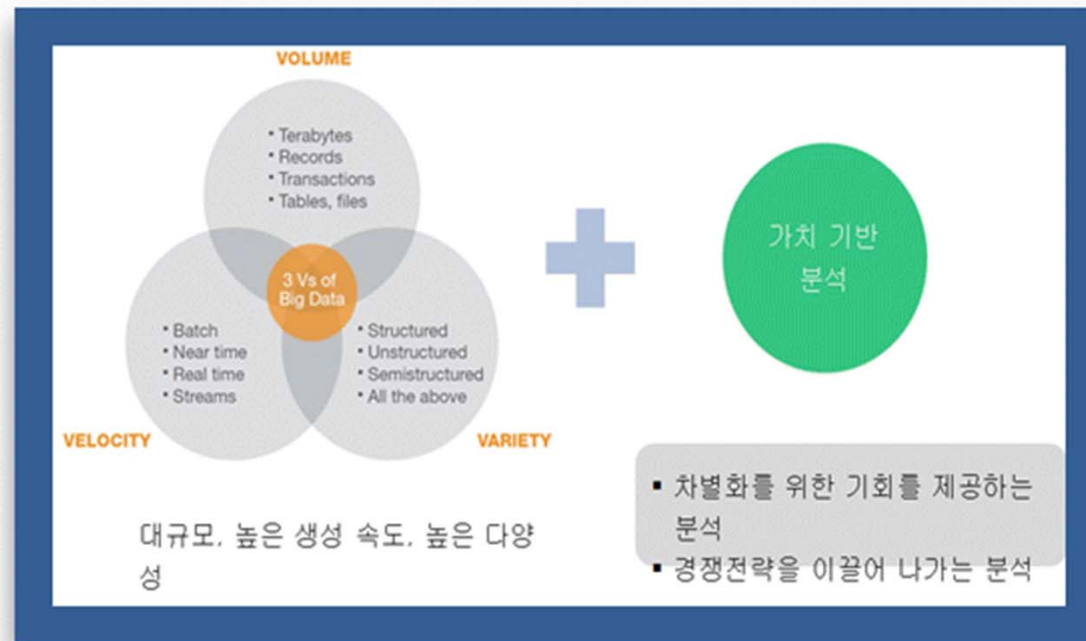
수집 방법에 따른 Data 분류

- **(1) 1차 자료**
- ㉠ 조사 목적에 적합한 자료를 얻기 위해서 조사자가 직접 수집한 자료★이다.
- **(2) 2차 자료**
- ㉠ 다른 조사목적으로 기존에 작성된 자료이다.

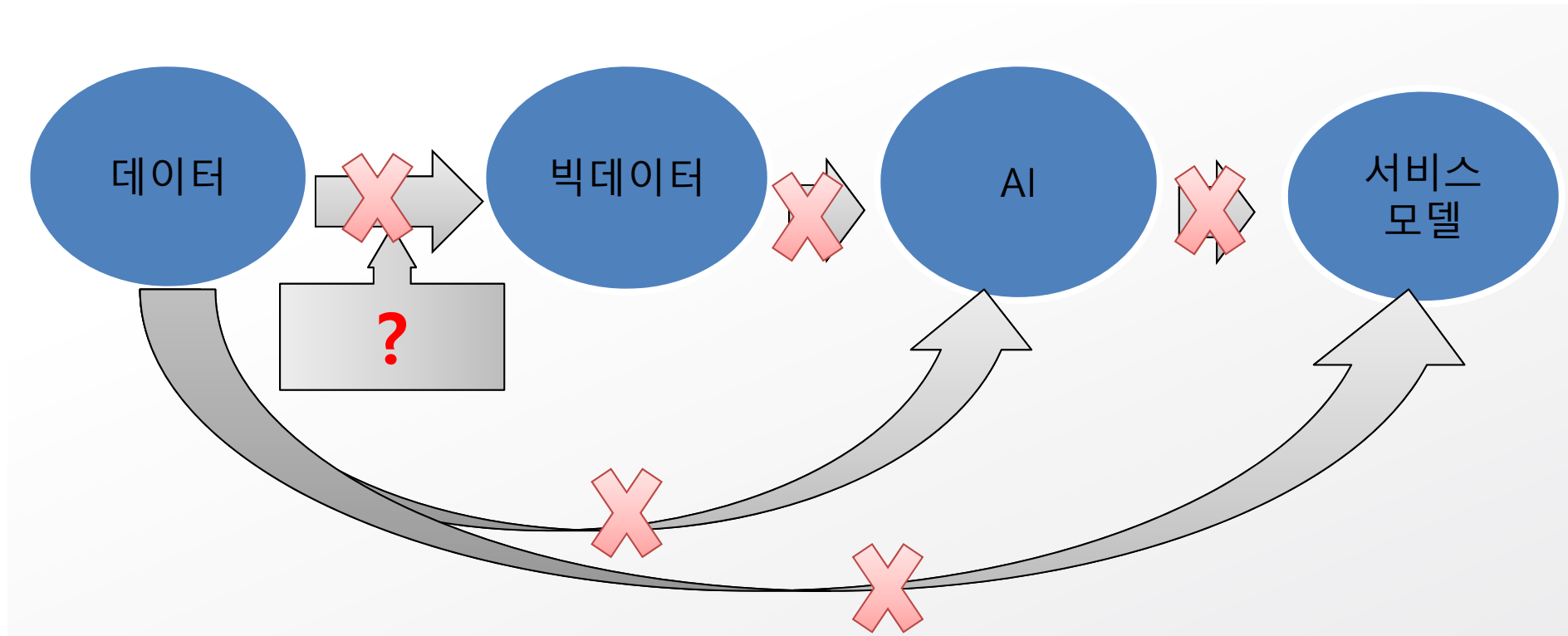
구조화 정도에 따른 Data 분류

구조화 정도에 따른 구분	설명
정형데이터	정형화된 속성으로 필드에 저장된 구조화된 데이터
준(반)정형데이터	필요시 관계형 데이터베이스에 쉽게 데이터를 저장 가능
구조화 가능데이터	구조화 할 수 있는 정보가 있어 구조화 가능한 데이터
비구조적 데이터	데이터 항목에 대해 사전에 정의된 설명이 없거나 구조화가 어려운 데이터

Big Data 란



데이터에서 AI 서비스 모델 개발 단계



빅데이터 구축은 데이터와 관련된 모든 팀(부서)의 협조가 필요.

데이터 범위 : 제공되는 서비스 관련 데이터

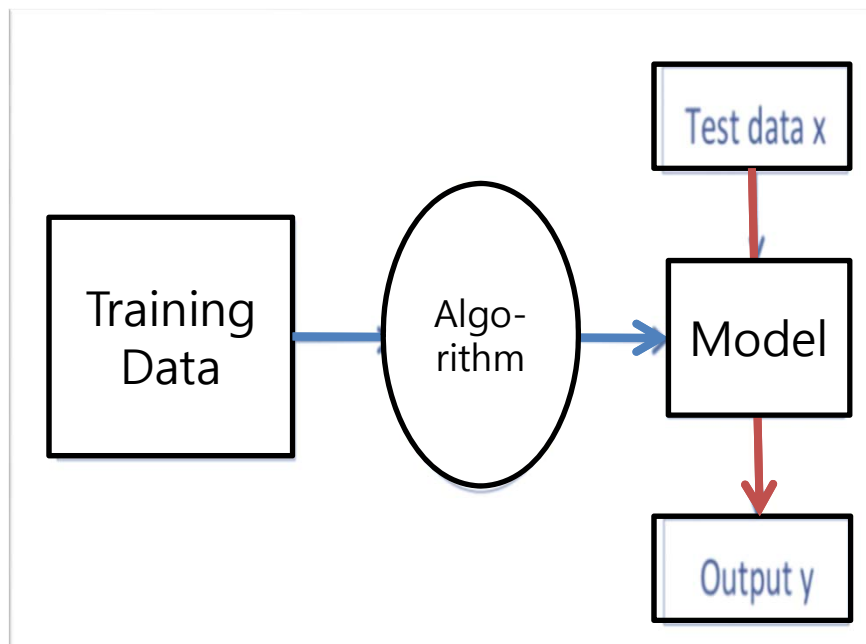
AI 서비스를 위해 각 팀(부서)에서 제공 가능 데이터 등

인공지능, 머신러닝, 딥러닝

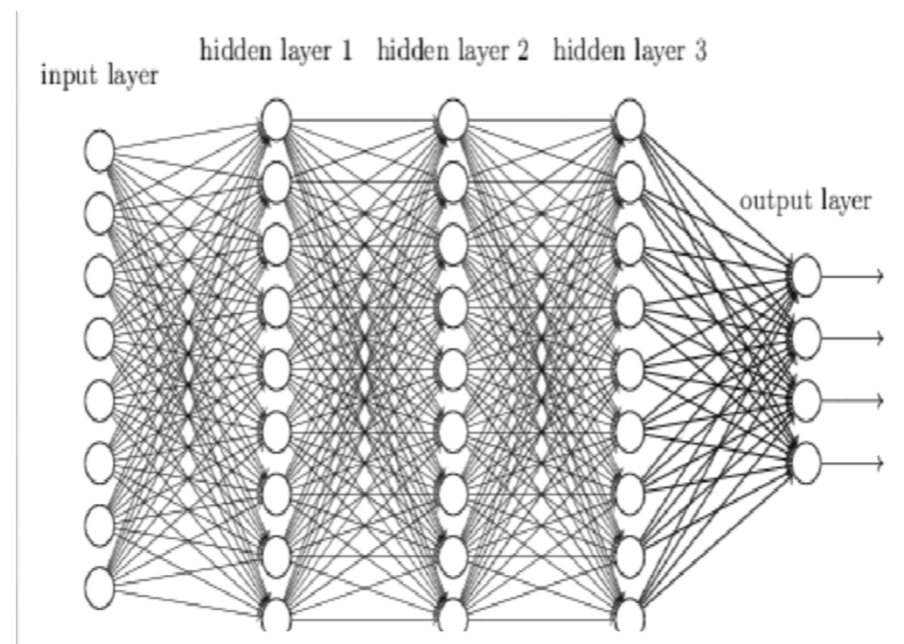
인공 지능(AI)

과거의 경험으로부터 자신(시스템)의 성능 향상, 즉 학습

머신 러닝



딥 러닝



목적치 유무에 따른 분류

: Target 변수에 따른 머신러닝의 분류

- 지도 학습 (supervised learning)
: 입력(문제)-출력(답)의 데이터로부터 출력을 결정할 수 있는 패턴추출
- 비지도 학습 (unsupervised learning, 자율 학습)
: 출력에 대한 정보가 없는 데이터로부터 필요한 패턴 추출
- 강화 학습 (reinforcement learning)
: 출력에 대한 정확한 정보를 제공하지는 않지만, 평가정보(reward)는 주어지는 문제에 대해 각 상태에서의 행동(action)을 결정