Introduction to Business Analytics: Data Science Overview

강필성
고려대학교 산업경영공학부
pilsung_kang@korea.ac.kr

AGENDA

01 빅데이터 분석 개요 및 주요 개념

02 데이터 과학 프로젝트 절차

03 기계 학습 방법론

04 제조업 활용 사례 I: 가상 계측 모델 개발

데이터 기반 의사결정

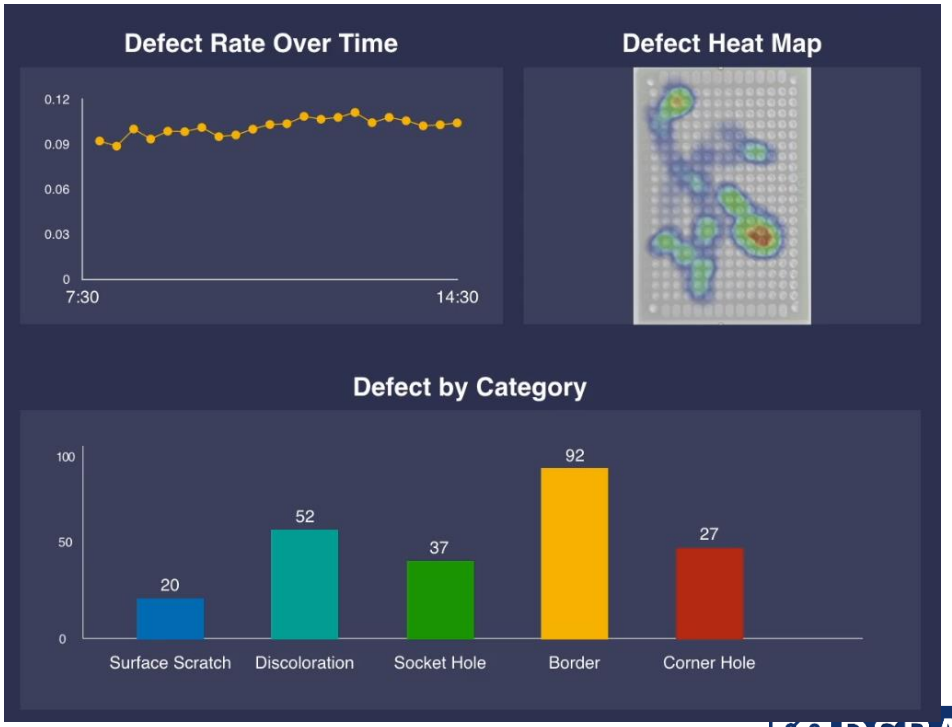
- 우리는 당신이 무엇을 구매할 지 이미 알고있다

데이터 기반 의사결정

- 우리가 알고 싶은 것

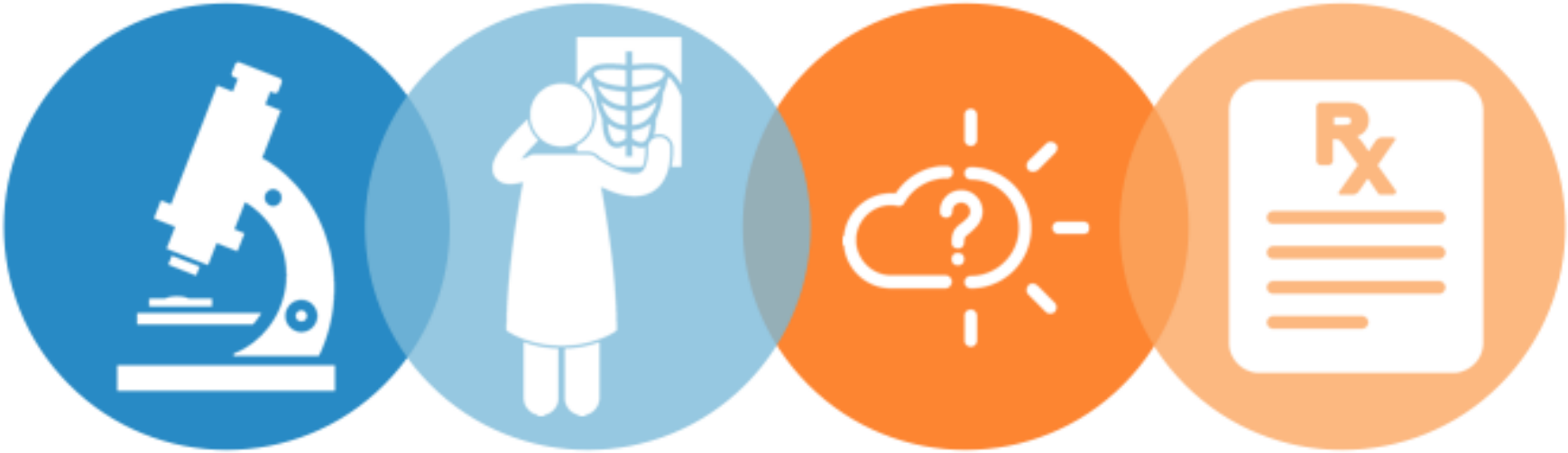
데이터 기반 의사결정

- 좀 더 구체적으로 제조업에서 무엇을 할 수 있을까?
  - ✓ **Landing.ai**
    - 인공지능 분야의 세계적 권위자인 Andrew Ng 교수가 인공지능의 제조업 적용을 목표로 세운 스타트업 (대만 폭스콘과 제휴)
    - 제품 이미지를 바탕으로 불량 판정 및 불량 의심 영역 판독



데이터 기반 의사결정

- 네 가지 유형의 Analytics



Descriptive
Explains what happened.

Diagnostic
Explains why it happened.

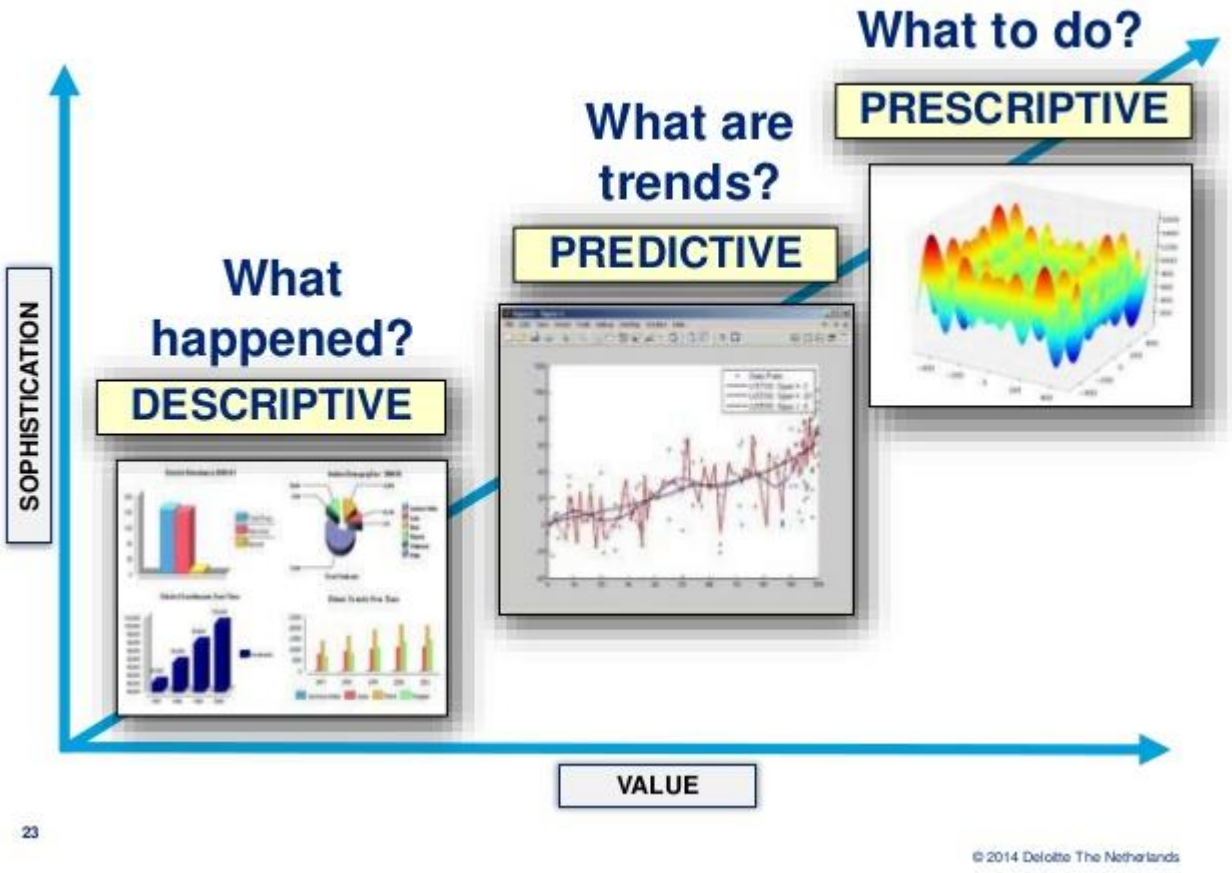
Predictive
Forecasts what might happen.

Prescriptive
Recommends an action based on the forecast.

https://www.analyticsinsight.net/four-types-of-business-analytics-to-know/

데이터 기반 의사결정

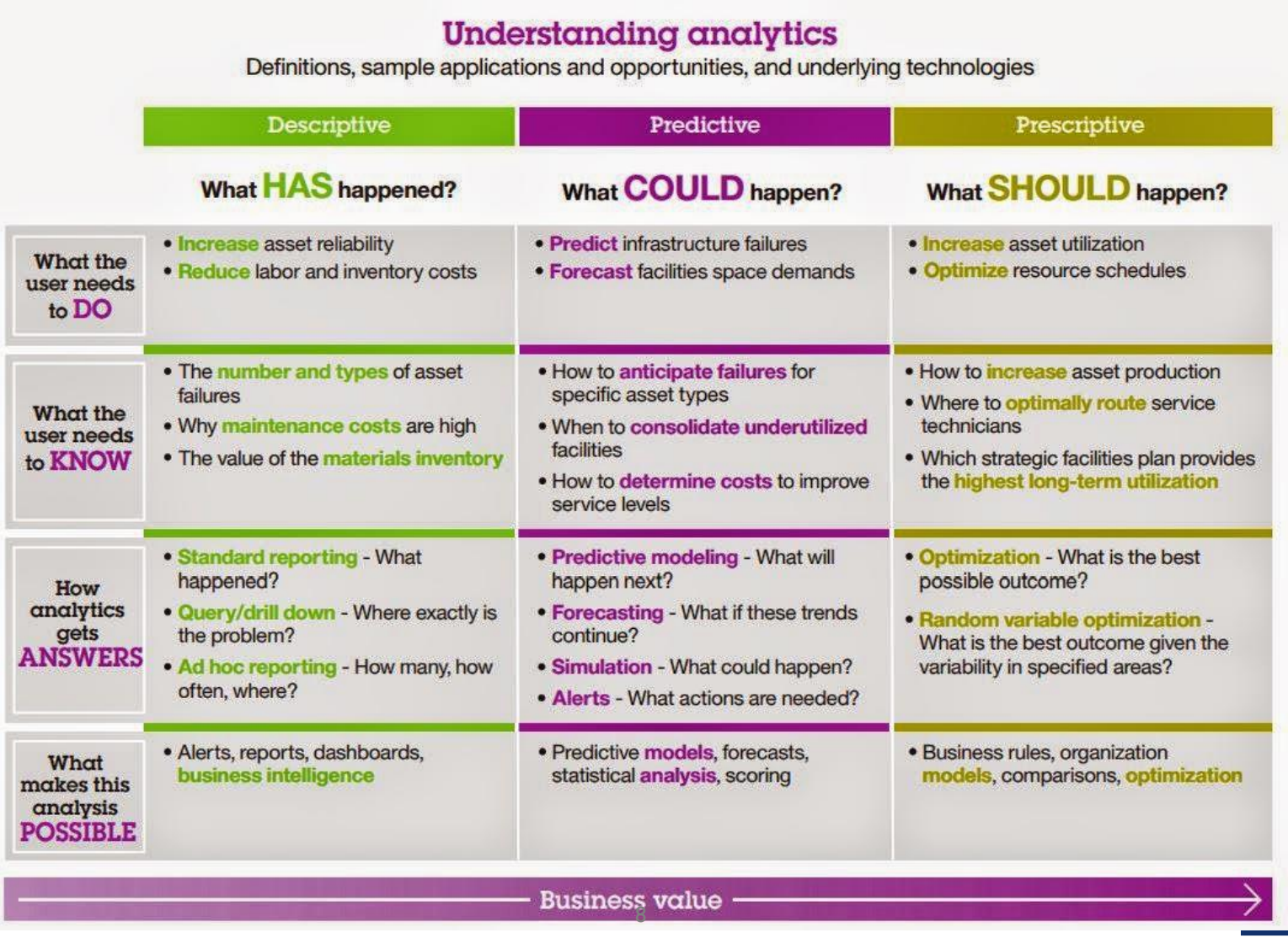
- 세 가지 유형의 Analytics



http://forums.bsdinsight.com/threads/descriptive-predictive-and-prescriptive-analytics-explained.41558/

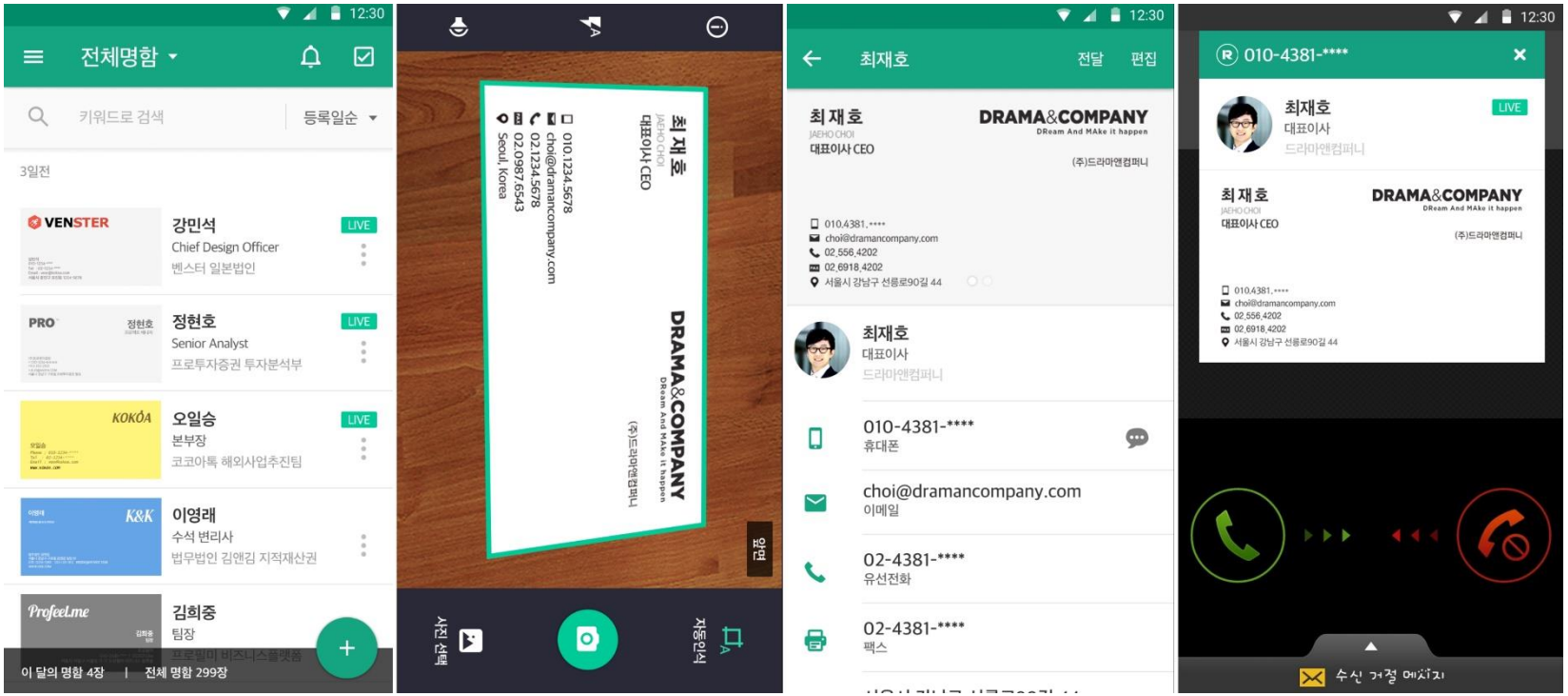
데이터 기반 의사결정

- 세 가지 유형의 Analytics



데이터 기반 의사결정: 데이터의 중요성

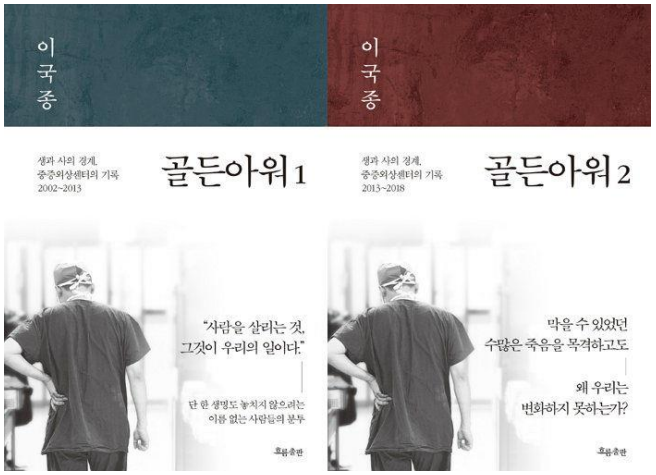
- 데이터는 수단인가? 아니면 목적인가?
  - ✓데이터를 얻기 위해 제품/서비스를 판매한다?



  - ✓결국 2017년 12월 21일 라인플러스와 네이버가 인수함

데이터 기반 의사결정: 데이터의 중요성

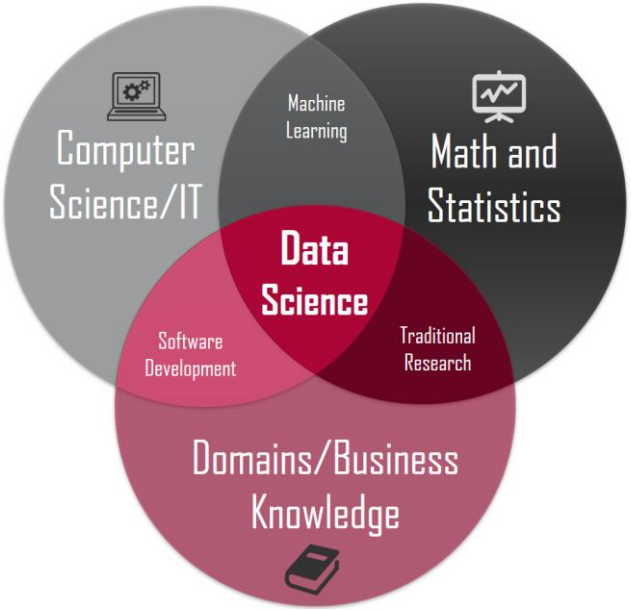
- 멋있는 알고리즘도 중요하지만 그 전에 먼저 제대로 된 데이터를 수집하자

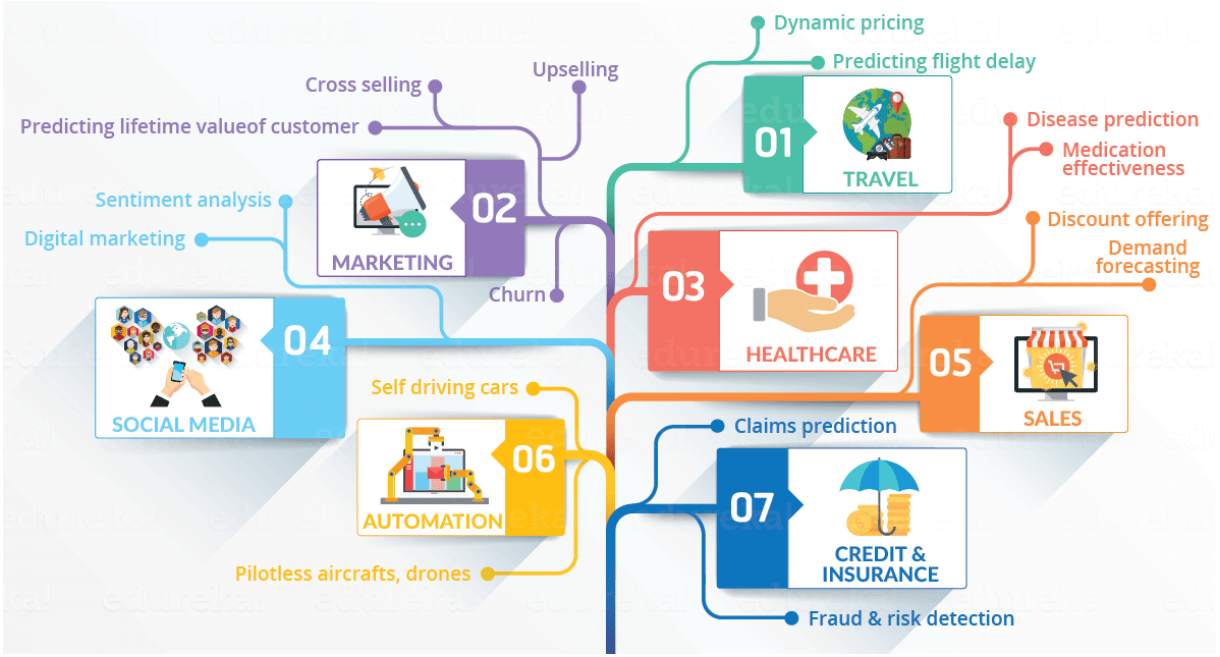


"문제는 한국 사회에서 시스템이 필요하다고 지시를 내릴 사람은 많은데 **전통적으로 '노가다'를 뛸 사람은 없다**는 겁니다. 이런 일은 남이 해야 하는 거라 생각하죠. 아니면 남이 했다가 자기한테 해가 되면 안 되니까 오만 가지 이유를 대서 이런 일은 하면 안 된다고 하고, 이런 일이 의료계에서만 있는 줄 알았어요. 사회 전반이 바뀌지 않으면 이 문제는 나아지지 않아요"라고 했다.

데이터 과학: 정의

- 데이터 과학이란?
  - ✓다양한 학제간 학문이 융합되어 데이터 기반 의사결정 및 문제 해결을 목적으로 하는 학문





데이터 과학: 연역법 vs. 귀납법

규칙:A속성의 사람들은 인사를 하고 B속성의 사람들은 악수를 한다

A와 B는 무엇일까?





데이터 과학: 연역법 vs. 귀납법

아시아계 사람들은 인사를 하고 백인들은 악수를 한다.





데이터 과학: 연역법 vs. 귀납법

아시아계 사람들은 인사를 하고 백인들은 악수를 한다.





데이터 과학: 연역법 vs. 귀납법

같은 색상의 옷을 입은 사람들은 인사를 하고,
다른 색상의 옷을 입은 사람들은 악수를 한다.





데이터 과학: 연역법 vs. 귀납법

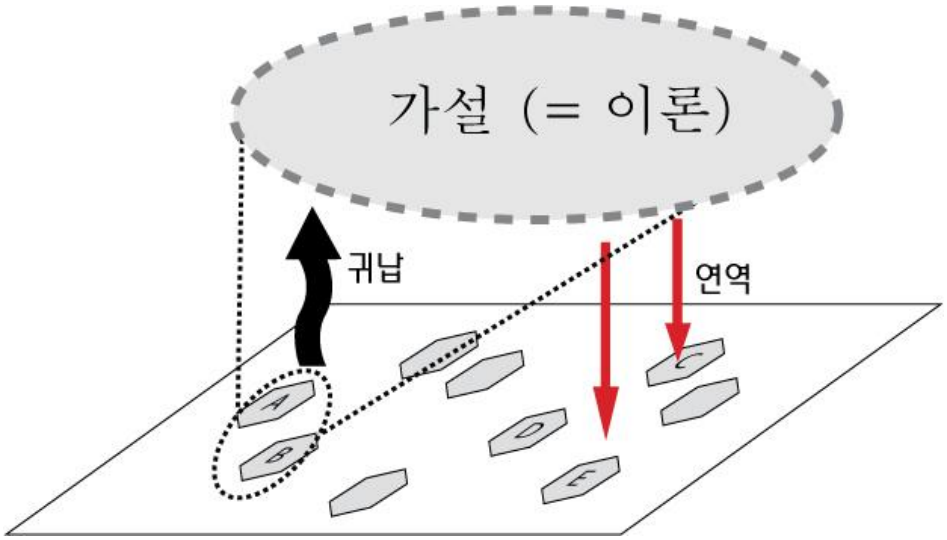
같은 색상의 옷을 입은 사람들은 인사를 하고,
다른 색상의 옷을 입은 사람들은 악수를 한다.

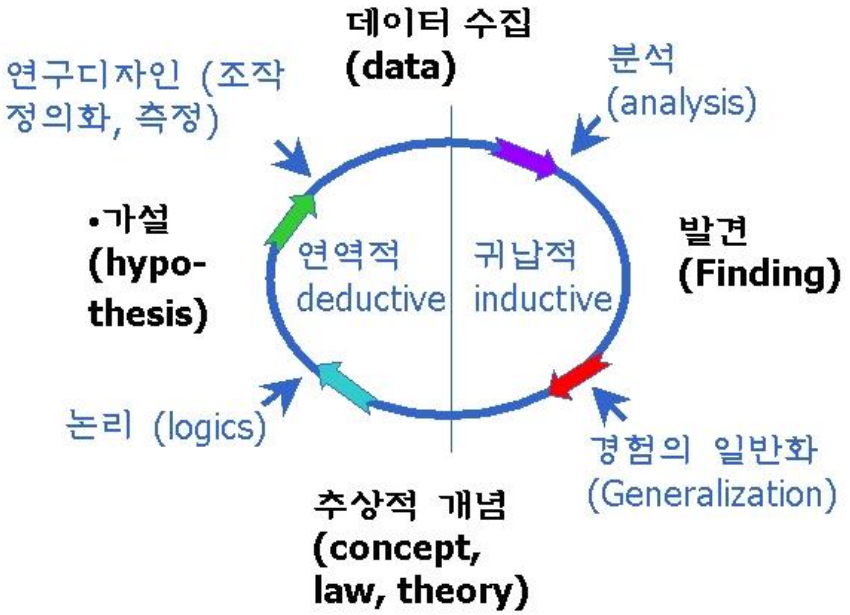




데이터 과학: 연역법 vs. 귀납법

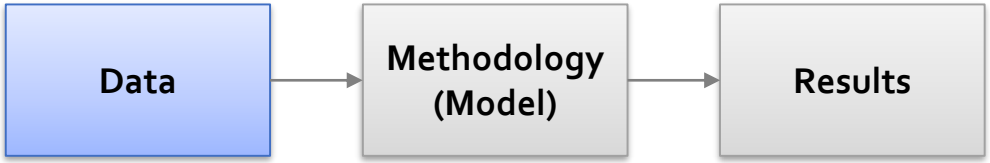
- **연역법**
  - ✓ 일반적 사실이나 원리를 전제로 하여 개별적인 특수한 사실이나 원리를 결론으로 이끌어 내는 추리 방법 (예: 삼단논법)
- **귀납법**
  - ✓ 여러 가지의 관찰된 사실들을 바탕으로 이들의 기저에 깔려 있는 일반적인 원리를 추론해 내는 방법





데이터 과학 주요 개념: 빅데이터

- 빅데이터(Big Data)



✓데이터베이스 규모에 초점을 맞춘 정의 (McKinsey, 2011)

- 일반적인 데이터베이스 SW가 저장, 관리, 분석할 수 있는 범위를 초과하는 규모의 데이터

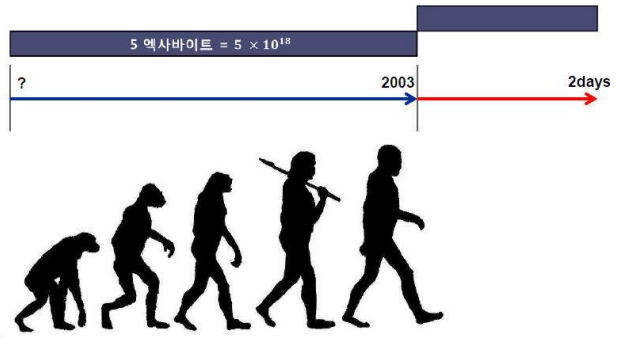
✓업무 수행에 초점을 맞춘 정의 (IDC, 2011)

- 다양한 종류의 대규모 데이터로부터 저렴한 비용으로 가치를 추출하고 초고속 수집, 발굴, 분석을 지원하도록 고안된 차세대 기술 및 아키텍처

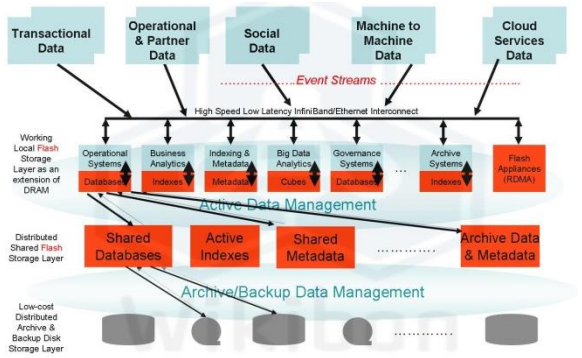
데이터 과학 주요 개념: 빅데이터

- 빅데이터의 4V
  - ✓빅데이터의 특징은 방대한 양(Volume), 빠른 데이터 생성 및 처리 속도(Velocity), 다양한 형태(Variety) 및 데이터에 내재된 잠재 가치(Value)로 정의됨

Volume



Velocity



Variety

Value

데이터 과학 주요 개념: 빅데이터

- 빅데이터의 특징
  - ✓복잡하고 고도화된 분석 방법론이 아닌 데이터 그 자체로서 가치를 가짐

VS

데이터 과학 주요 개념: 빅데이터

- 빅데이터의 특징

  ✓복잡하고 고도화된 분석 방법론이 아닌 데이터 그 자체로서 가치를 가짐

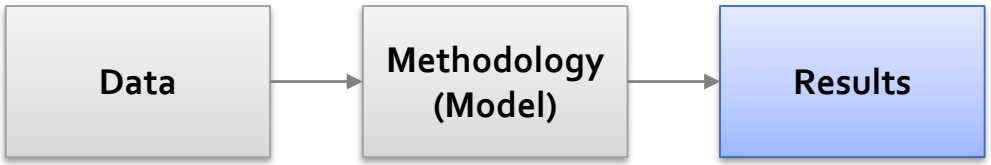
- 데이터에 의한 정량적 유동인구 분포도 작성
- 서울시를 1km 반경의 1,250개 헥사셀 단위로 구분
- KT 휴대전화 이력 데이터로 심야시간 (0시~5시) 통화량 분석
- 유동인구 기반 노선 최적화 및 배차간격 조정

데이터 과학 주요 개념: 빅데이터

- 빅데이터의 특징
  - ✓복잡하고 고도화된 분석 방법론이 아닌 데이터 그 자체로서 가치를 가짐
    - 지멘스 암베르크 공장 사례

데이터 과학 주요 개념: 데이터 마이닝

- 데이터 마이닝: Data Mining
  - ✓ 대량의 데이터로부터 의미있는 규칙이나 패턴을 추출하는 일련의 활동



- ✓ Extracting useful information from large datasets. (Hand et al., 2001)
- ✓ The process of exploration and analysis, by automatic or semi-automatic means, of large quantities of data in order to discover meaningful patterns and rules. (Berry and Linoff, 1997, 2000)
- ✓ The process of discovering meaningful new correlations, patterns and trends by sifting through large amount data stored in repositories, using pattern recognition technologies as well as statistical and mathematical techniques. (Gartner Group, 2004)

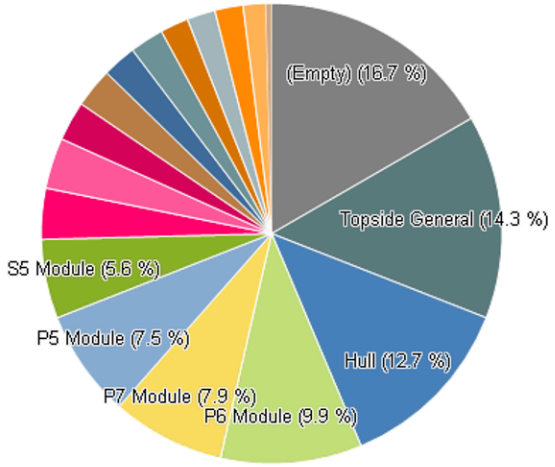
데이터 과학 주요 개념: 데이터 마이닝

- 데이터 마이닝: Data Mining
  - ✓대량의 데이터로부터 의미있는 규칙이나 패턴을 추출하는 일련의 활동

"파이프(pipe)가 흔들리니(shake), 지주(support)를 추가(add)하라"

언제, 어디서?

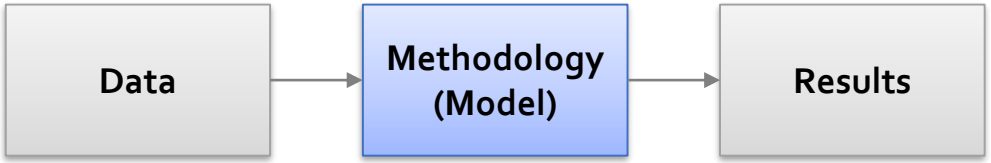
"공사 중반, Topside General, Hull, P5,6,7 Module 등에서 주로 발생한다"

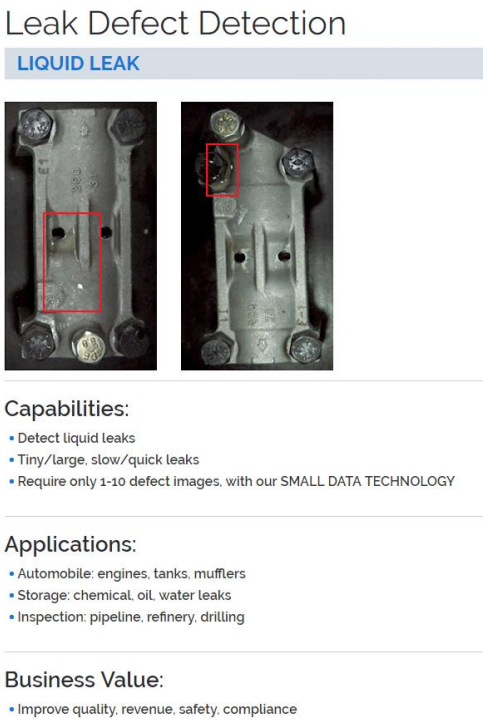


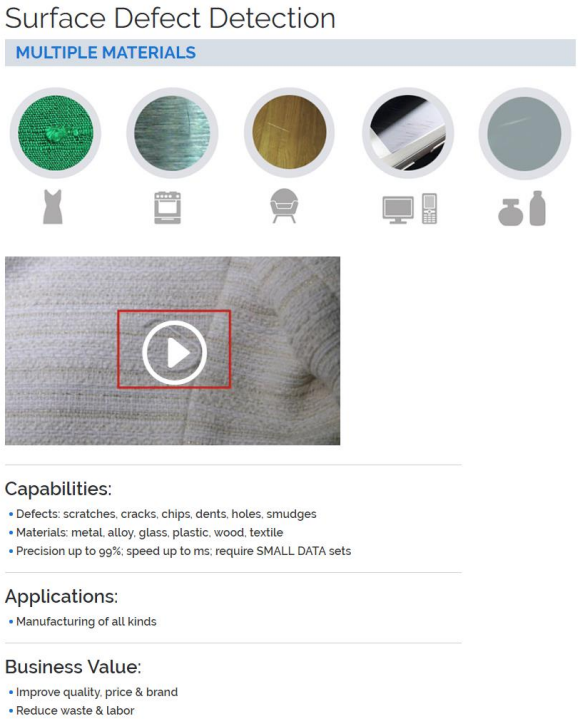
발생하는 장소

데이터 과학 주요 개념: 기계 학습

- 기계 학습: Machine Learning
  - ✓ 특정한 과업Task을 달성하기 위해 경험Experience이 축적될수록 과업 수행의 성능Performance이 향상되는 컴퓨터 프로그램 또는 에이전트를 개발하는 것 – Mitchell (1997)

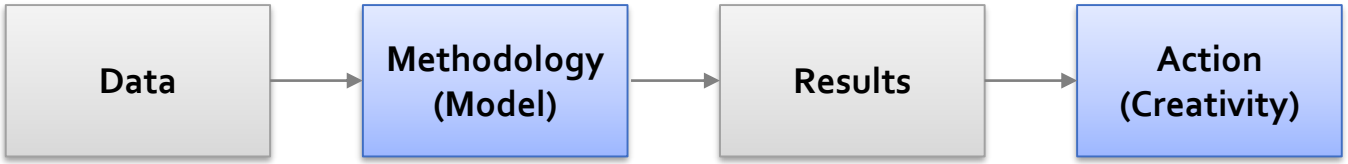






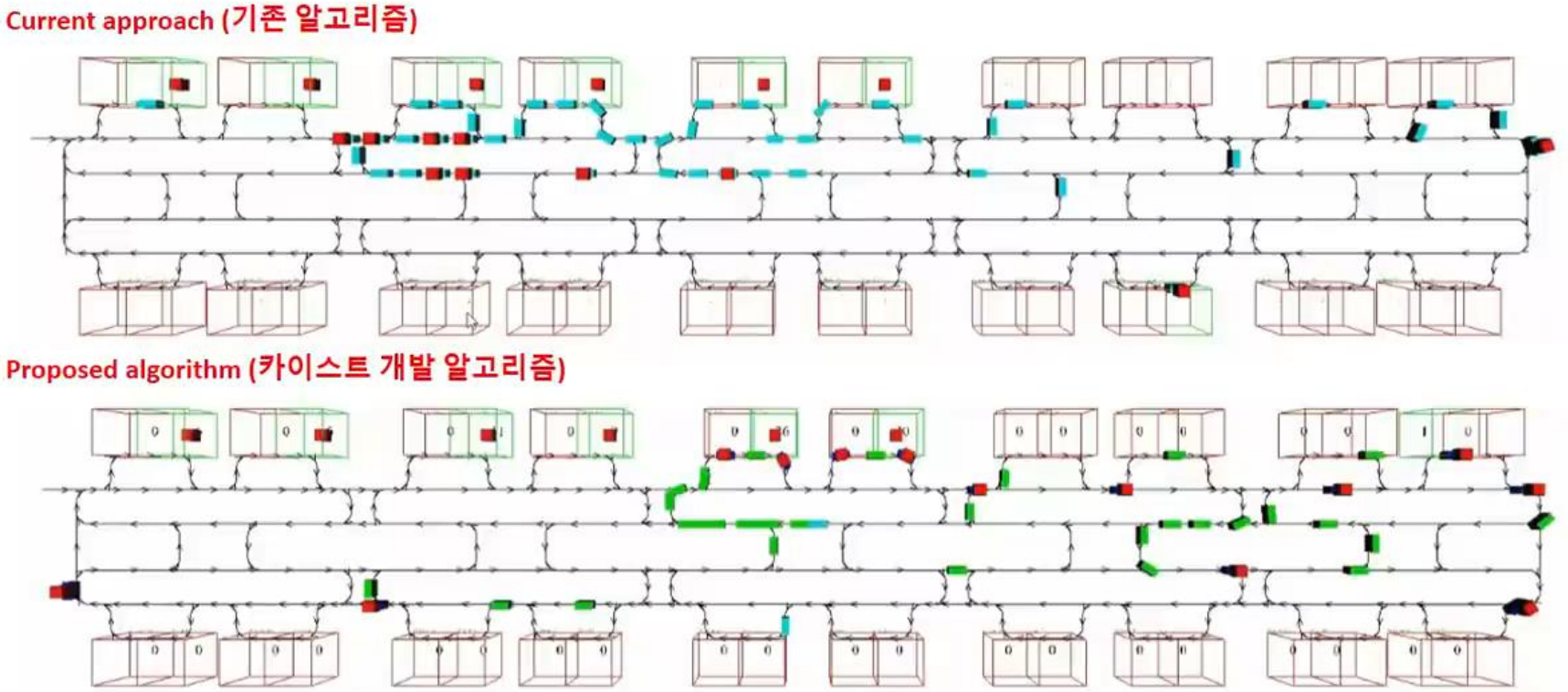
데이터 과학 주요 개념: 인공 지능

- 인공 지능: Artificial Intelligence
  - ✓환경을 인지하여 보상이 최대화되는 지능적인 행위를 할 수 있는 컴퓨터 소프트웨어



데이터 과학 주요 개념: 인공 지능

- 강화학습을 이용한 물류 최적화

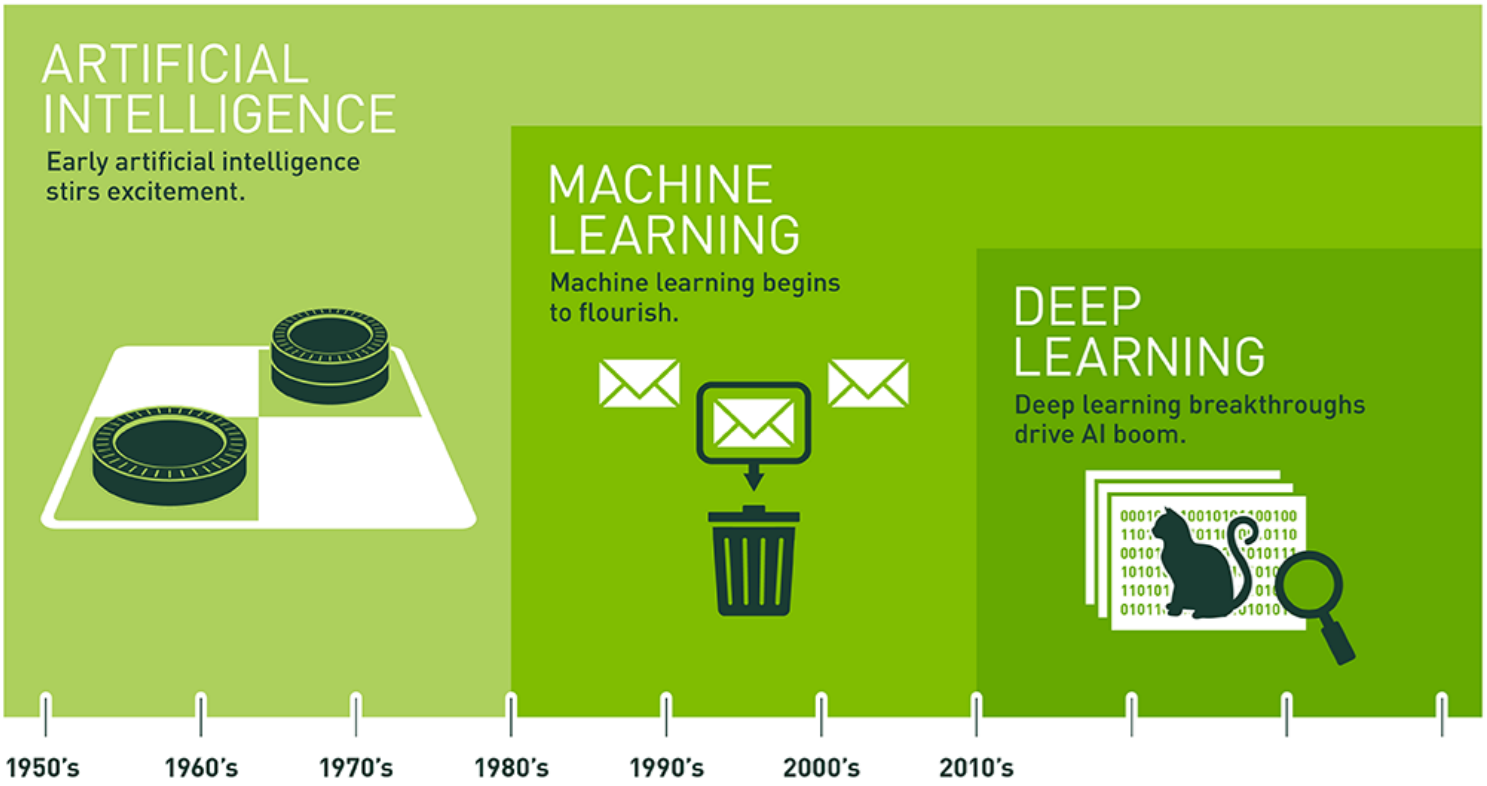


Illhoe Hwang, Sang Pyo Hong, Young Jae Jang, Sunil Kim and In-Ho Moon, "System Design and Development of the Q-Learning Based Overhead Hoist Transport (OHT) for Semiconductor Fabs," International Symposium on Semiconductor Initiatives, 2018

Information: http://sdm.kaist.ac.kr

데이터 과학 주요 개념: 인공 지능

- 인공지능 vs. 기계학습 vs. 딥러닝
  - ✓인공지능이 가장 상위 개념이며 딥러닝은 기계학습의 한 종류임

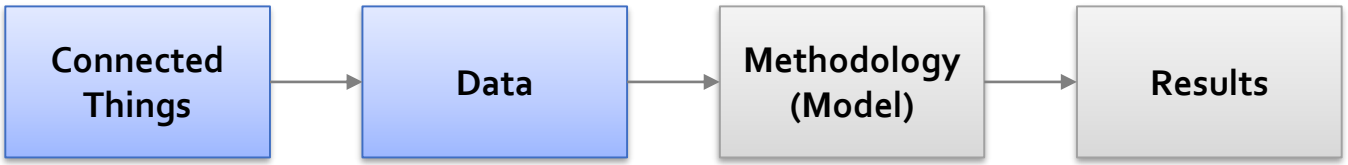


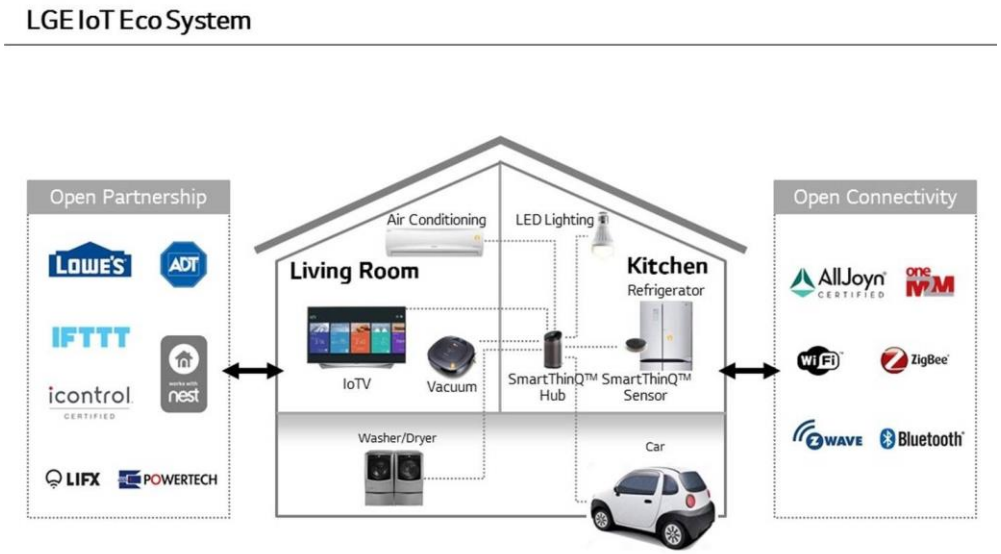
Since an early flush of optimism in the 1950s, smaller subsets of artificial intelligence – first machine learning, then deep learning, a subset of machine learning – have created ever larger disruptions.

출처: http://blogs.nvidia.com/blog/2016/07/29/whats-difference-artificial-intelligence-machine-learning-deep-learning-ai/

데이터 과학 주요 개념: 사물 인터넷

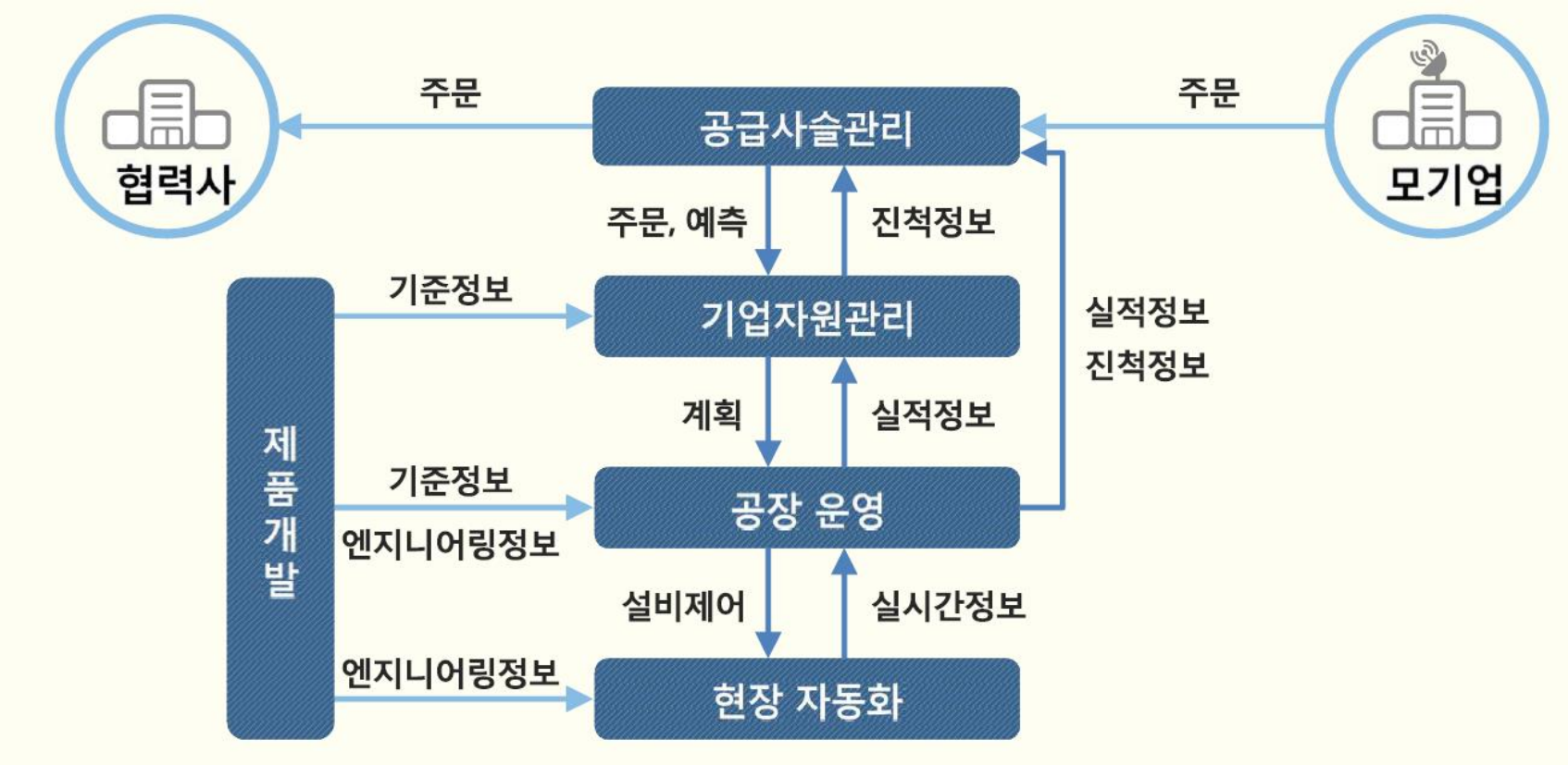
- 사물 인터넷
  - ✓센서 및 소프트웨어가 내장된 물리적 개체들이 연결어 각 개체들간의 통신, 데이터 교환, 컨트롤 등을 지원하는 네트워크 체계





데이터 과학 주요 개념: 사물 인터넷

- 사물 인터넷: Internet of Things
  - ✓4차 산업 혁명과 스마트 공장(Smart Factory)의 핵심 구성 요소




ANY
questions?